

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/64942 A1

(51) International Patent Classification⁷: **C12Q 1/00**,
1/70, 1/68, G01N 33/53

Concord, MA 01742 (US). **KUIMELIS, Robert, G.**
[US/US]; 21 Malbert Road, Brighton, MA 02135 (US).

(21) International Application Number: PCT/US01/06414

(74) Agent: **ELBING, Karen, L.**; Clark & Elbing LLP, 176
Federal Street, Boston, MA 02110-2214 (US).

(22) International Filing Date: 28 February 2001 (28.02.2001)

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/515,260 29 February 2000 (29.02.2000) US

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 09/515,260 (CIP)
Filed on 29 February 2000 (29.02.2000)

(71) Applicant (*for all designated States except US*): **PHY-
LOS, INC.** [US/US]; 128 Spring Street, Lexington, MA
02421 (US).

Published:

- with international search report
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **LIPOVSEK, Dasa**
[SI/US]; 45 Sunset Road, Cambridge, MA 02138 (US).
WAGNER, Richard, W. [US/US]; 1007 Lowell Road,

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS AND OTHER BINDING PROTEINS

(57) Abstract: Disclosed herein are proteins that include a fibronectin type III domain having at least one randomized loop. Also disclosed herein are nucleic acids encoding such proteins and the use of such proteins in diagnostic methods and in methods for evolving novel compound-binding species and their ligands.

WO 01/64942 A1

PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS
AND OTHER BINDING PROTEINS

5

Background of the Invention

This invention relates to protein scaffolds useful, for example, for the generation of products having novel binding characteristics.

Proteins having relatively defined three-dimensional structures, commonly referred to as protein scaffolds, may be used as reagents for the design of engineered products. These scaffolds typically contain one or more regions which are amenable to specific or random sequence variation, and such sequence randomization is often carried out to produce libraries of proteins from which desired products may be selected. One particular area in which such scaffolds are useful is the field of antibody design.

A number of previous approaches to the manipulation of the mammalian immune system to obtain reagents or drugs have been attempted. These have included injecting animals with antigens of interest to obtain mixtures of polyclonal antibodies reactive against specific antigens, production of monoclonal antibodies in hybridoma cell culture (Koehler and Milstein, Nature 256:495, 1975), modification of existing monoclonal antibodies to obtain new or optimized recognition properties, creation of novel antibody fragments with desirable binding characteristics, and randomization of single chain antibodies (created by connecting the variable regions of the heavy and light chains of antibody molecules with a flexible peptide linker) followed by selection for antigen binding by phage display (Clackson et al., Nature 352:624, 1991).

In addition, several non-immunoglobulin protein scaffolds have been proposed for obtaining proteins with novel binding properties. For example, a "minibody" scaffold, which is related to the immunoglobulin fold, has been designed by deleting three beta strands from a heavy chain variable domain of a monoclonal antibody (Tramontano et al., J. Mol. Recognit. 7:9, 1994). This protein includes 61 residues and can be used to present two hypervariable loops. These two loops have been randomized and products selected for antigen binding, but thus far the framework appears to have somewhat limited utility due to solubility problems. Another framework used to display loops has been tendamistat, a 74 residue, six-strand beta sheet sandwich held together by two disulfide bonds (McConnell and Hoess, J. Mol. Biol. 250:460, 1995). This scaffold includes three loops, but, to date, only two of these loops have been examined for randomization potential.

Other proteins have been tested as frameworks and have been used to display randomized residues on alpha helical surfaces (Nord et al., Nat. Biotechnol. 15:772, 1997; Nord et al., Protein Eng. 8:601, 1995), loops between alpha helices in alpha helix bundles (Ku and Schultz, Proc. Natl. Acad. Sci. USA 92:6552, 1995), and loops constrained by disulfide bridges, such as those of the small protease inhibitors (Markland et al., Biochemistry 35:8045, 1996; Markland et al., Biochemistry 35:8058, 1996; Rottgen and Collins, Gene 164:243, 1995; Wang et al., J. Biol. Chem. 270:12250, 1995).

Summary of the Invention

The present invention provides a new family of proteins capable of evolving to bind any compound of interest. These proteins, which make use of a fibronectin or fibronectin-like scaffold, function in a manner characteristic of natural or engineered antibodies (that is, polyclonal, monoclonal, or

single-chain antibodies) and, in addition, possess structural advantages. Specifically, the structure of these antibody mimics has been designed for optimal folding, stability, and solubility, even under conditions which normally lead to the loss of structure and function in antibodies.

5 These antibody mimics may be utilized for the purpose of designing proteins which are capable of binding to virtually any compound (for example, any protein) of interest. In particular, the fibronectin-based molecules described herein may be used as scaffolds which are subjected to directed evolution designed to randomize one or more of the three fibronectin loops
10 which are analogous to the complementarity-determining regions (CDRs) of an antibody variable region. Such a directed evolution approach results in the production of antibody-like molecules with high affinities for antigens of interest. In addition, the scaffolds described herein may be used to display defined exposed loops (for example, loops previously randomized and selected
15 on the basis of antigen binding) in order to direct the evolution of molecules that bind to such introduced loops. A selection of this type may be carried out to identify recognition molecules for any individual CDR-like loop or, alternatively, for the recognition of two or all three CDR-like loops combined into a non-linear epitope.

20 Accordingly, the present invention features a protein that includes a fibronectin type III domain having at least one randomized loop, the protein being characterized by its ability to bind to a compound that is not bound by the corresponding naturally-occurring fibronectin.

 In preferred embodiments, the fibronectin type III domain is a
25 mammalian (for example, a human) fibronectin type III domain; and the protein includes the tenth module of the fibronectin type III (¹⁰F_n3) domain. In such proteins, compound binding is preferably mediated by either one, two, or

three ¹⁰F_n3 loops. In other preferred embodiments, the second loop of ¹⁰F_n3 may be extended in length relative to the naturally-occurring module, or the ¹⁰F_n3 may lack an integrin-binding motif. In these molecules, the integrin-binding motif may be replaced by an amino acid sequence in which a basic
5 amino acid-neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the integrin-binding motif; one preferred sequence is serine-glycine-glutamate. In another preferred embodiment, the fibronectin type III domain-containing proteins of the invention lack disulfide bonds.

10 Any of the fibronectin type III domain-containing proteins described herein may be formulated as part of a fusion protein (for example, a fusion protein which further includes an immunoglobulin F_c domain, a complement protein, a toxin protein, or an albumin protein). In addition, any of the fibronectin type III domain proteins may be covalently bound to a nucleic acid
15 (for example, an RNA), and the nucleic acid may encode the protein. Moreover, the protein may be a multimer, or, particularly if it lacks an integrin-binding motif, it may be formulated in a physiologically-acceptable carrier.

The present invention also features proteins that include a fibronectin type III domain having at least one mutation in a β -sheet sequence
20 which changes the scaffold structure. Again, these proteins are characterized by their ability to bind to compounds that are not bound by the corresponding naturally-occurring fibronectin.

In addition, any of the fibronectin scaffolds of the invention may be immobilized on a solid support (for example, a bead or chip), and these
25 scaffolds may be arranged in any configuration on the solid support, including an array.

In a related aspect, the invention further features nucleic acids encoding any of the proteins of the invention. In preferred embodiments, the nucleic acid is DNA or RNA.

In another related aspect, the invention also features a method for
5 generating a protein which includes a fibronectin type III domain and which is pharmaceutically acceptable to a mammal, involving removing the integrin-binding domain of said fibronectin type III domain. This method may be applied to any of the fibronectin type III domain-containing proteins described above and is particularly useful for generating proteins for human therapeutic
10 applications. The invention also features such fibronectin type III domain-containing proteins which lack integrin-binding domains.

In yet other related aspects, the invention features screening methods which may be used to obtain or evolve randomized fibronectin type III proteins capable of binding to compounds of interest, or to obtain or evolve compounds
15 (for example, proteins) capable of binding to a particular protein containing a randomized fibronectin type III motif. In addition, the invention features screening procedures which combine these two methods, in any order, to obtain either compounds or proteins of interest.

In particular, the first screening method, useful for the isolation or
20 identification of randomized proteins of interest, involves: (a) contacting the compound with a candidate protein, the candidate protein including a fibronectin type III domain having at least one randomized loop, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) obtaining, from the complex, the protein which binds to the
25 compound.

The second screening method, for isolating or identifying a compound which binds to a protein having a randomized fibronectin type III domain, involves: (a) contacting the protein with a candidate compound, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) obtaining, from the complex, the compound which binds to the protein.

In preferred embodiments, the methods further involve either randomizing at least one loop of the fibronectin type III domain of the protein obtained in step (b) and repeating steps (a) and (b) using the further randomized protein, or modifying the compound obtained in step (b) and repeating steps (a) and (b) using the further modified compound. In addition, the compound is preferably a protein, and the fibronectin type III domain is preferably a mammalian (for example, a human) fibronectin type III domain. In other preferred embodiments, the protein includes the tenth module of the fibronectin type III domain (¹⁰F_n3), and binding is mediated by one, two, or three ¹⁰F_n3 loops. In addition, the second loop of ¹⁰F_n3 may be extended in length relative to the naturally-occurring module, or ¹⁰F_n3 may lack an integrin-binding motif. Again, as described above, the integrin-binding motif may be replaced by an amino acid sequence in which a basic amino acid-neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the integrin-binding motif; one preferred sequence is serine-glycine-glutamate.

The selection methods described herein may be carried out using any fibronectin type III domain-containing protein. For example, the fibronectin type III domain-containing protein may lack disulfide bonds, or may be formulated as part of a fusion protein (for example, a fusion protein which further includes an immunoglobulin F_c domain, a complement protein, a toxin

protein, or an albumin protein). In addition, selections may be carried out using the fibronectin type III domain proteins covalently bound to nucleic acids (for example, RNAs or any nucleic acid which encodes the protein). Moreover, the selections may be carried out using fibronectin domain-
5 containing protein multimers.

Preferably, the selections involve the immobilization of the binding target on a solid support. Preferred solid supports include columns (for example, affinity columns, such as agarose columns) or microchips.

In addition, the invention features diagnostic methods which employ
10 the fibronectin scaffold proteins of the invention. Such diagnostic methods may be carried out on a sample (for example, a biological sample) to detect one analyte or to simultaneously detect many different analytes in the sample. The method may employ any of the scaffold molecules described herein. Preferably, the method involves (a) contacting the sample with a protein which
. 15 binds to the compound analyte and which includes a fibronectin type III domain having at least one randomized loop, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) detecting the complex, and therefore the compound in the sample.

In preferred embodiments, the protein is immobilized on a solid
20 support (for example, a chip or bead) and may be immobilized as part of an array. The protein may be covalently bound to a nucleic acid, preferably, a nucleic acid, such as RNA, that encodes the protein. In addition, the compound is often a protein, but may also be any other analyte in a sample. Detection may be accomplished by any standard technique including, without
25 limitation, radiography, fluorescence detection, mass spectroscopy, or surface plasmon resonance.

As used herein, by “fibronectin type III domain” is meant a domain having 7 or 8 beta strands which are distributed between two beta sheets, which themselves pack against each other to form the core of the protein, and further containing loops which connect the beta strands to each other and are solvent exposed. There are at least three such loops at each edge of the beta sheet sandwich, where the edge is the boundary of the protein perpendicular to the direction of the beta strands. Preferably, a fibronectin type III domain includes a sequence which exhibits at least 30% amino acid identity, and preferably at least 50% amino acid identity, to the sequence encoding the structure of the ¹⁰F_n3 domain referred to as “1ttg” (ID = “1ttg” (one ttg)) available from the Protein Data Base. Sequence identity referred to in this definition is determined by the Homology program, available from Molecular Simulation (San Diego, CA). The invention further includes polymers of ¹⁰F_n3-related molecules, which are an extension of the use of the monomer structure, whether or not the subunits of the polyprotein are identical or different in sequence.

By “naturally occurring fibronectin” is meant any fibronectin protein that is encoded by a living organism.

By “randomized” is meant including one or more amino acid alterations relative to a template sequence.

By a “protein” is meant any sequence of two or more amino acids, regardless of length, post-translation modification, or function. “Protein” and “peptide” are used interchangeably herein.

By “RNA” is meant a sequence of two or more covalently bonded, naturally occurring or modified ribonucleotides. One example of a modified RNA included within this term is phosphorothioate RNA.

By "DNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified deoxyribonucleotides.

By a "nucleic acid" is meant any two or more covalently bonded nucleotides or nucleotide analogs or derivatives. As used herein, this term
5 includes, without limitation, DNA, RNA, and PNA.

By "pharmaceutically acceptable" is meant a compound or protein that may be administered to an animal (for example, a mammal) without significant adverse medical consequences.

By "physiologically acceptable carrier" is meant a carrier which does
10 not have a significant detrimental impact on the treated host and which retains the therapeutic properties of the compound with which it is administered. One exemplary physiologically acceptable carrier is physiological saline. Other physiologically acceptable carriers and their formulations are known to one skilled in the art and are described, for example, in Remington's
15 Pharmaceutical Sciences, (18th edition), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA, incorporated herein by reference.

By "selecting" is meant substantially partitioning a molecule from other molecules in a population. As used herein, a "selecting" step provides at least a 2-fold, preferably, a 30-fold, more preferably, a 100-fold, and, most
20 preferably, a 1000-fold enrichment of a desired molecule relative to undesired molecules in a population following the selection step. A selection step may be repeated any number of times, and different types of selection steps may be combined in a given approach.

By "binding partner," as used herein, is meant any molecule which
25 has a specific, covalent or non-covalent affinity for a portion of a desired compound (for example, protein) of interest. Examples of binding partners include, without limitation, members of antigen/antibody pairs,

protein/inhibitor pairs, receptor/ligand pairs (for example cell surface receptor/ligand pairs, such as hormone receptor/peptide hormone pairs), enzyme/substrate pairs (for example, kinase/substrate pairs), lectin/carbohydrate pairs, oligomeric or heterooligomeric protein aggregates, DNA binding protein/DNA binding site pairs, RNA/protein pairs, and nucleic acid duplexes, heteroduplexes, or ligated strands, as well as any molecule which is capable of forming one or more covalent or non-covalent bonds (for example, disulfide bonds) with any portion of another molecule (for example, a compound or protein).

10 By a "solid support" is meant, without limitation, any column (or column material), bead, test tube, microtiter dish, solid particle (for example, agarose or sepharose), microchip (for example, silicon, silicon-glass, or gold chip), or membrane (for example, the membrane of a liposome or vesicle) to which a fibronectin scaffold or an affinity complex may be bound, either
15 directly or indirectly (for example, through other binding partner intermediates such as other antibodies or Protein A), or in which a fibronectin scaffold or an affinity complex may be embedded (for example, through a receptor or channel).

The present invention provides a number of advantages. For
20 example, as described in more detail below, the present antibody mimics exhibit improved biophysical properties, such as stability under reducing conditions and solubility at high concentrations. In addition, these molecules may be readily expressed and folded in prokaryotic systems, such as E. coli, in eukaryotic systems, such as yeast, and in in vitro translation systems, such as
25 the rabbit reticulocyte lysate system. Moreover, these molecules are extremely amenable to affinity maturation techniques involving multiple cycles of selection, including in vitro selection using RNA-protein fusion technology

(Roberts and Szostak, Proc. Natl. Acad. Sci USA 94:12297, 1997; Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al.

WO98/31700), phage display (see, for example, Smith and Petrenko, Chem. Rev. 97:317, 1997), and yeast display systems (see, for example, Boder and

5 Wittrup, Nature Biotech. 15:553, 1997).

Other features and advantages of the present invention will be apparent from the following detailed description thereof, and from the claims.

Brief Description of the Drawings

FIGURE 1 is a photograph showing a comparison between the
10 structures of antibody heavy chain variable regions from camel (dark blue) and llama (light blue), in each of two orientations.

FIGURE 2 is a photograph showing a comparison between the structures of the camel antibody heavy chain variable region (dark blue), the llama antibody heavy chain variable region (light blue), and a fibronectin type
15 III module number 10 (¹⁰F_n3) (yellow).

FIGURE 3 is a photograph showing a fibronectin type III module number 10 (¹⁰F_n3), with the loops corresponding to the antigen-binding loops in IgG heavy chains highlighted in red.

FIGURE 4 is a graph illustrating a sequence alignment between a
20 fibronectin type III protein domain and related protein domains.

FIGURE 5 is a photograph showing the structural similarities between a ¹⁰F_n3 domain and 15 related proteins, including fibronectins, tenascins, collagens, and undulin. In this photograph, the regions are labeled as follows: constant, dark blue; conserved, light blue; neutral, white; variable,
25 red; and RGB integrin-binding motif (variable), yellow.

FIGURE 6 is a photograph showing space filling models of fibronectin III modules 9 and 10, in each of two different orientations. The two modules and the integrin binding loop (RGB) are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

FIGURE 7 is a photograph showing space filling models of fibronectin III modules 7-10, in each of three different orientations. The four modules are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

FIGURE 8 is a photograph illustrating the formation, under different salt conditions, of RNA-protein fusions which include fibronectin type III domains.

FIGURE 9 is a series of photographs illustrating the selection of fibronectin type III domain-containing RNA-protein fusions, as measured by PCR signal analysis.

FIGURE 10 is a graph illustrating an increase in the percent TNF- α binding during the selections described herein, as well as a comparison between RNA-protein fusion and free protein selections.

FIGURE 11 is a series of schematic representations showing IgG, $^{10}\text{Fn3}$, $\text{Fn-CH}_1\text{-CH}_2\text{-CH}_3$, and $\text{Fn-CH}_2\text{-CH}_3$ (clockwise from top left).

FIGURE 12 is a photograph showing a molecular model of $\text{Fn-CH}_1\text{-CH}_2\text{-CH}_3$ based on known three-dimensional structures of IgG (X-ray crystallography) and $^{10}\text{Fn3}$ (NMR and X-ray crystallography).

FIGURE 13 is a graph showing the time course of an exemplary $^{10}\text{Fn3}$ -based nucleic acid-protein fusion selection of TNF- α binders. The proportion of nucleic acid-protein fusion pool (open diamonds) and free

protein pool (open circles) that bound to TNF- α -Sephadrose, and the proportion of free protein pool (full circles) that bound to underivatized Sephadrose, are shown.

FIGURES 14 and 15 are graphs illustrating TNF- α binding by TNF- α Fn-binders. In particular, these figures show mass spectra data obtained from a $^{10}\text{Fn3}$ fusion chip and non-fusion chip, respectively.

FIGURES 16 and 17 are the phosphorimage and fluorescence scan, respectively, of a $^{10}\text{Fn3}$ array, illustrating TNF- α binding.

Detailed Description

The novel antibody mimics described herein have been designed to be superior both to antibody-derived fragments and to non-antibody frameworks, for example, those frameworks described above.

The major advantage of these antibody mimics over antibody fragments is structural. These scaffolds are derived from whole, stable, and soluble structural modules found in human body fluid proteins. Consequently, they exhibit better folding and thermostability properties than antibody fragments, whose creation involves the removal of parts of the antibody native fold, often exposing amino acid residues that, in an intact antibody, would be buried in a hydrophobic environment, such as an interface between variable and constant domains. Exposure of such hydrophobic residues to solvent increases the likelihood of aggregation.

In addition, the antibody mimics described herein have no disulfide bonds, which have been reported to retard or prevent proper folding of antibody fragments under certain conditions. Since the present scaffolds do not rely on disulfides for native fold stability, they are stable under reducing conditions, unlike antibodies and their fragments which unravel upon disulfide

bond breakdown.

Moreover, these fibronectin-based scaffolds provide the functional advantages of antibody molecules. In particular, despite the fact that the ¹⁰F_n3 module is not an immunoglobulin, its overall fold is close to that of the variable region of the IgG heavy chain (Figure 2), making it possible to display the three fibronectin loops analogous to CDRs in relative orientations similar to those of native antibodies. Because of this structure, the present antibody mimics possess antigen binding properties that are similar in nature and affinity to those of antibodies, and a loop randomization and shuffling strategy may be employed in vitro that is similar to the process of affinity maturation of antibodies in vivo.

There are now described below exemplary fibronectin-based scaffolds and their use for identifying, selecting, and evolving novel binding proteins as well as their target ligands. These examples are provided for the purpose of illustrating, and not limiting, the invention.

¹⁰F_n3 Structural Motif

The antibody mimics of the present invention are based on the structure of a fibronectin module of type III (F_n3), a common domain found in mammalian blood and structural proteins. This domain occurs more than 400 times in the protein sequence database and has been estimated to occur in 2% of the proteins sequenced to date, including fibronectins, tensin, intracellular cytoskeletal proteins, and prokaryotic enzymes (Bork and Doolittle, Proc. Natl. Acad. Sci. USA 89:8990, 1992; Bork et al., Nature Biotech. 15:553, 1997; Meinke et al., J. Bacteriol. 175:1910, 1993; Watanabe et al., J. Biol. Chem. 265:15659, 1990). In particular, these scaffolds include, as templates, the tenth module of human F_n3 (¹⁰F_n3), which comprises 94 amino acid residues.

The overall fold of this domain is closely related to that of the smallest functional antibody fragment, the variable region of the heavy chain, which comprises the entire antigen recognition unit in camel and llama IgG (Figure 1, 2). The major differences between camel and llama domains and the ¹⁰F_n3 domain are that (i) ¹⁰F_n3 has fewer beta strands (seven vs. nine) and (ii) the two beta sheets packed against each other are connected by a disulfide bridge in the camel and llama domains, but not in ¹⁰F_n3.

The three loops of ¹⁰F_n3 corresponding to the antigen-binding loops of the IgG heavy chain run between amino acid residues 21-31, 51-56, and 76-88 (Figure 3). The length of the first and the third loop, 11 and 12 residues, respectively, fall within the range of the corresponding antigen-recognition loops found in antibody heavy chains, that is, 10-12 and 3-25 residues, respectively. Accordingly, once randomized and selected for high antigen affinity, these two loops make contacts with antigens equivalent to the contacts of the corresponding loops in antibodies.

In contrast, the second loop of ¹⁰F_n3 is only 6 residues long, whereas the corresponding loop in antibody heavy chains ranges from 16-19 residues. To optimize antigen binding, therefore, the second loop of ¹⁰F_n3 is preferably extended by 10-13 residues (in addition to being randomized) to obtain the greatest possible flexibility and affinity in antigen binding. Indeed, in general, the lengths as well as the sequences of the CDR-like loops of the antibody mimics may be randomized during in vitro or in vivo affinity maturation (as described in more detail below).

The tenth human fibronectin type III domain, ¹⁰F_n3, refolds rapidly even at low temperature; its backbone conformation has been recovered within 1 second at 5°C. Thermodynamic stability of ¹⁰F_n3 is high ($\Delta G_U = 24$ kJ/mol = 5.7 kcal/mol), correlating with its high melting temperature of 110°C.

One of the physiological roles of $^{10}\text{Fn3}$ is as a subunit of fibronectin, a glycoprotein that exists in a soluble form in body fluids and in an insoluble form in the extracellular matrix (Dickinson et al., J. Mol. Biol. 236:1079, 1994). A fibronectin monomer of 220-250 kD contains 12 type I modules, two
5 type II modules, and 17 fibronectin type III modules (Potts and Campbell, Curr. Opin. Cell Biol. 6:648, 1994). Different type III modules are involved in the binding of fibronectin to integrins, heparin, and chondroitin sulfate. $^{10}\text{Fn3}$ was found to mediate cell adhesion through an integrin-binding Arg-Gly-Asp (RGD) motif on one of its exposed loops. Similar RGD motifs have been
10 shown to be involved in integrin binding by other proteins, such as fibrinogen, von Willebrand factor, and vitronectin (Hynes et al., Cell 69:11, 1992). No other matrix- or cell-binding roles have been described for $^{10}\text{Fn3}$.

The observation that $^{10}\text{Fn3}$ has only slightly more adhesive activity than a short peptide containing RGD is consistent with the conclusion that the
15 cell-binding activity of $^{10}\text{Fn3}$ is localized in the RGD peptide rather than distributed throughout the $^{10}\text{Fn3}$ structure (Baron et al., Biochemistry 31:2068, 1992). The fact that $^{10}\text{Fn3}$ without the RGD motif is unlikely to bind to other plasma proteins or extracellular matrix makes $^{10}\text{Fn3}$ a useful scaffold to replace antibodies. In addition, the presence of $^{10}\text{Fn3}$ in natural fibrinogen in the
20 bloodstream suggests that $^{10}\text{Fn3}$ itself is unlikely to be immunogenic in the organism of origin.

In addition, we have determined that the $^{10}\text{Fn3}$ framework possesses exposed loop sequences tolerant of randomization, facilitating the generation of diverse pools of antibody mimics. This determination was made by
25 examining the flexibility of the $^{10}\text{Fn3}$ sequence. In particular, the human $^{10}\text{Fn3}$ sequence was aligned with the sequences of fibronectins from other sources as well as sequences of related proteins (Figure 4), and the results of this

alignment were mapped onto the three-dimensional structure of the human ¹⁰F_n3 domain (Figure 5). This alignment revealed that the majority of conserved residues are found in the core of the beta sheet sandwich, whereas the highly variable residues are located along the edges of the beta sheets, including the N- and C-termini, on the solvent-accessible faces of both beta sheets, and on three solvent-accessible loops that serve as the hypervariable loops for affinity maturation of the antibody mimics. In view of these results, the randomization of these three loops are unlikely to have an adverse effect on the overall fold or stability of the ¹⁰F_n3 framework itself.

For the human ¹⁰F_n3 sequence, this analysis indicates that, at a minimum, amino acids 1-9, 44-50, 61-54, 82-94 (edges of beta sheets); 19, 21, 30-46 (even), 79-65 (odd) (solvent-accessible faces of both beta sheets); 21-31, 51-56, 76-88 (CDR-like solvent-accessible loops); and 14-16 and 36-45 (other solvent-accessible loops and beta turns) may be randomized to evolve new or improved compound-binding proteins. In addition, as discussed above, alterations in the lengths of one or more solvent exposed loops may also be included in such directed evolution methods. Alternatively, changes in the β -sheet sequences may also be used to evolve new proteins. These mutations change the scaffold and thereby indirectly alter loop structure(s). If this approach is taken, mutations should not saturate the sequence, but rather few mutations should be introduced. Preferably, no more than 10 amino acid changes, and, more preferably, no more than 3 amino acid changes should be introduced to the β -sheet sequences by this approach.

Fibronectin Fusions

The antibody mimics described herein may be fused to other protein domains. For example, these mimics may be integrated with the human

immune response by fusing the constant region of an IgG (F_c) with a $^{10}\text{Fn3}$ module, preferably through the C-terminus of $^{10}\text{Fn3}$. The F_c in such a $^{10}\text{Fn3}$ - F_c fusion molecule activates the complement component of the immune response and increases the therapeutic value of the antibody mimic. Similarly, a fusion
5 between $^{10}\text{Fn3}$ and a complement protein, such as C1q, may be used to target cells, and a fusion between $^{10}\text{Fn3}$ and a toxin may be used to specifically destroy cells that carry a particular antigen. In addition, $^{10}\text{Fn3}$ in any form may be fused with albumin to increase its half-life in the bloodstream and its tissue penetration. Any of these fusions may be generated by standard techniques,
10 for example, by expression of the fusion protein from a recombinant fusion gene constructed using publically available gene sequences.

Fibronectin Scaffold Multimers

In addition to fibronectin monomers, any of the fibronectin constructs described herein may be generated as dimers or multimers of
15 $^{10}\text{Fn3}$ -based antibody mimics as a means to increase the valency and thus the avidity of antigen binding. Such multimers may be generated through covalent binding between individual $^{10}\text{Fn3}$ modules, for example, by imitating the natural $^8\text{Fn3}$ - $^9\text{Fn3}$ - $^{10}\text{Fn3}$ C-to-N-terminus binding or by imitating antibody dimers that are held together through their constant regions. A $^{10}\text{Fn3}$ - F_c
20 construct may be exploited to design dimers of the general scheme of $^{10}\text{Fn3}$ - $F_c::F_c$ - $^{10}\text{Fn3}$. The bonds engineered into the $F_c::F_c$ interface may be covalent or non-covalent. In addition, dimerizing or multimerizing partners other than F_c can be used in $^{10}\text{Fn3}$ hybrids to create such higher order structures.

In particular examples, covalently bonded multimers may be generated by constructing fusion genes that encode the multimer or, alternatively, by engineering codons for cysteine residues into monomer sequences and allowing disulfide bond formation to occur between the expression products. Non-covalently bonded multimers may also be generated by a variety of techniques. These include the introduction, into monomer sequences, of codons corresponding to positively and/or negatively charged residues and allowing interactions between these residues in the expression products (and therefore between the monomers) to occur. This approach may be simplified by taking advantage of charged residues naturally present in a monomer subunit, for example, the negatively charged residues of fibronectin. Another means for generating non-covalently bonded antibody mimics is to introduce, into the monomer gene (for example, at the amino- or carboxy-termini), the coding sequences for proteins or protein domains known to interact. Such proteins or protein domains include coil-coil motifs, leucine zipper motifs, and any of the numerous protein subunits (or fragments thereof) known to direct formation of dimers or higher order multimers.

Fibronectin-Like Molecules

Although ¹⁰F_n3 represents a preferred scaffold for the generation of antibody mimics, other molecules may be substituted for ¹⁰F_n3 in the molecules described herein. These include, without limitation, human fibronectin modules ¹F_n3-⁹F_n3 and ¹¹F_n3-¹⁷F_n3 as well as related F_n3 modules from non-human animals and prokaryotes. In addition, F_n3 modules from other proteins with sequence homology to ¹⁰F_n3, such as tenascins and undulins, may also be used. Modules from different organisms and parent proteins may be most appropriate for different applications; for example, in

designing an antibody mimic, it may be most desirable to generate that protein from a fibronectin or fibronectin-like molecule native to the organism for which a therapeutic or diagnostic molecule is intended.

Directed Evolution of Scaffold-Based Binding Proteins

5 The antibody mimics described herein may be used in any technique for evolving new or improved binding proteins. In one particular example, the target of binding is immobilized on a solid support, such as a column resin or microtiter plate well, and the target contacted with a library of candidate scaffold-based binding proteins. Such a library may consist of ¹⁰F_n3 clones
10 constructed from the wild type ¹⁰F_n3 scaffold through randomization of the sequence and/or the length of the ¹⁰F_n3 CDR-like loops. If desired, this library may be an RNA-protein fusion library generated, for example, by the techniques described in Szostak et al., U.S.S.N. 09/007,005 and 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci.
15 USA (1997) vol. 94, p. 12297-12302. Alternatively, it may be a DNA-protein library (for example, as described in Lohse, DNA-Protein Fusions and Uses Thereof, U.S.S.N. 60/110,549, U.S.S.N. 09/459,190, and US 99/28472). The fusion library is incubated with the immobilized target, the support is washed to remove non-specific binders, and the tightest binders are eluted under very
20 stringent conditions and subjected to PCR to recover the sequence information or to create a new library of binders which may be used to repeat the selection process, with or without further mutagenesis of the sequence. A number of rounds of selection may be performed until binders of sufficient affinity for the antigen are obtained.

In one particular example, the ¹⁰F_n3 scaffold may be used as the selection target. For example, if a protein is required that binds a specific peptide sequence presented in a ten residue loop, a single ¹⁰F_n3 clone is constructed in which one of its loops has been set to the length of ten and to
5 the desired sequence. The new clone is expressed *in vivo* and purified, and then immobilized on a solid support. An RNA-protein fusion library based on an appropriate scaffold is then allowed to interact with the support, which is then washed, and desired molecules eluted and re-selected as described above.

Similarly, the ¹⁰F_n3 scaffold may be used to find natural proteins
10 that interact with the peptide sequence displayed in a ¹⁰F_n3 loop. The ¹⁰F_n3 protein is immobilized as described above, and an RNA-protein fusion library is screened for binders to the displayed loop. The binders are enriched through multiple rounds of selection and identified by DNA sequencing.

In addition, in the above approaches, although RNA-protein libraries
15 represent exemplary libraries for directed evolution, any type of scaffold-based library may be used in the selection methods of the invention.

Use

The antibody mimics described herein may be evolved to bind any antigen of interest. These proteins have thermodynamic properties superior to
20 those of natural antibodies and can be evolved rapidly in vitro. Accordingly, these antibody mimics may be employed in place of antibodies in all areas in which antibodies are used, including in the research, therapeutic, and diagnostic fields. In addition, because these scaffolds possess solubility and stability properties superior to antibodies, the antibody mimics described
25 herein may also be used under conditions which would destroy or inactivate antibody molecules. Finally, because the scaffolds of the present invention

may be evolved to bind virtually any compound, these molecules provide completely novel binding proteins which also find use in the research, diagnostic, and therapeutic areas.

Experimental Results

5 Exemplary scaffold molecules described above were generated and tested, for example, in selection protocols, as follows.

Library construction

A complex library was constructed from three fragments, each of which contained one randomized area corresponding to a CDR-like loop. The
10 fragments were named BC, DE, and FG, based on the names of the CDR-H-like loops contained within them; in addition to ¹⁰F_n3 and a randomized sequence, each of the fragments contained stretches encoding an N-terminal His₆ domain or a C-terminal FLAG peptide tag. At each junction between two fragments (i.e., between the BC and DE fragments or between the
15 DE and FG fragments), each DNA fragment contained recognition sequences for the *E*arI Type IIS restriction endonuclease. This restriction enzyme allowed the splicing together of adjacent fragments while removing all foreign, non-¹⁰F_n3, sequences. It also allows for a recombination-like mixing of the three ¹⁰F_n3 fragments between cycles of mutagenesis and selection.

20 Each fragment was assembled from two overlapping oligonucleotides, which were first annealed, then extended to form the double-stranded DNA form of the fragment. The oligonucleotides that were used to construct and process the three fragments are listed below; the "Top" and "Bottom" species for each fragment are the oligonucleotides that contained
25 the entire ¹⁰F_n3 encoding sequence. In these oligonucleotides designations,

“N” indicates A, T, C, or G; and “S” indicates C or G.

HfnLbcTop (His):

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA
TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT
5 GTT CCG AGG GAC CTG GAA GTT GTT GCT GCG ACC CCC ACC
AGC-3' (SEQ ID NO: 1)

HfnLbcTop (an alternative N-terminus):

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA
TTT ACA ATT ACA ATG GTT TCT GAT GTT CCG AGG GAC CTG
10 GAA GTT GTT GCT GCG ACC CCC ACC AGC-3' (SEQ ID NO: 2)

HFnLBCBot-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT
CCC TGT TTC TCC GTA AGT GAT CCT GTA ATA TCT (SNN)⁷ CCA
GCT GAT CAG TAG GCT GGT GGG GGT CGC AGC -3' (SEQ ID NO: 3)

15 HFnBC3'-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT
CCC TGT TTC TCC GTA AGT GAT CC-3' (SEQ ID NO: 4)

HFnLDETop:

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA
20 TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC ACA
GGA GGA AAT AGC CCT GTC C-3' (SEQ ID NO: 5)

HF_nLDEBot-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT
CGT ATA ATC AAC TCC AGG TTT AAG GCC GCT GAT GGT AGC
TGT (SNN)⁴ AGG CAC AGT GAA CTC CTG GAC AGG GCT ATT TCC
5 TCC TGT -3' (SEQ ID NO: 6)

HF_nDE3'-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT
CGT ATA ATC AAC TCC AGG TTT AAG G-3' (SEQ ID NO: 7)

HF_nLFGTop:

10 5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA
TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC TAT
ACC ATC ACT GTG TAT GCT GTC-3' (SEQ ID NO: 8)

HF_nLFGBot-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG
15 GTA ATT AAT GGA AAT TGG (SNN)¹⁰ AGT GAC AGC ATA CAC AGT
GAT GGT ATA -3' (SEQ ID NO: 9)

HF_nFG3'-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG
GTA ATT AAT GGA AAT TGG -3' (SEQ ID NO: 10)

20 T7Tmv (introduces T7 promoter and TMV untranslated region needed for in
vitro translation):

5'- GCG TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA

ATT ACA-3' (SEQ ID NO: 11)

ASAflag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC-3' (SEQ ID NO: 12)

- 5 Unispl-s (spint oligonucleotide used to ligate mRNA to the puromycin-containing linker, described by Roberts et al, 1997, supra):
5'-TTTTTTTTTNAGCGGATGC-3' (SEQ ID NO: 13)

A18---2PEG (DNA-puromycin linker):

5'-(A)18(PEG)2CCPur (SEQ ID NO: 14)

- 10 The pairs of oligonucleotides (500 pmol of each) were annealed in 100 μ L of 10 mM Tris 7.5, 50 mM NaCl for 10 minutes at 85°C, followed by a slow (0.5-1 hour) cooling to room temperature. The annealed fragments with single-stranded overhangs were then extended using 100 U Klenow (New England Biolabs, Beverly, MA) for each 100 μ L aliquot of annealed oligos,
15 and the buffer made of 838.5 μ l H₂O, 9 μ l 1 M Tris 7.5, 5 μ l 1M MgCl₂, 20 μ l 10 mM dNTPs, and 7.5 μ l 1M DTT. The extension reactions proceeded for 1 hour at 25°C.

- Next, each of the double-stranded fragments was transformed into a RNA-protein fusion (PROfusionTM) using the technique developed by Szostak
20 et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. Briefly, the fragments were transcribed using an Ambion in vitro transcription kit, MEGAshortscript (Ambion, Austin, TX), and the

resulting mRNA was gel-purified and ligated to a DNA-puromycin linker using DNA ligase. The mRNA-DNA-puromycin molecule was then translated using the Ambion rabbit reticulocyte lysate-based translation kit. The resulting mRNA-DNA-puromycin-protein PROfusion™ was purified using Oligo(dT) 5 cellulose, and a complementary DNA strand was synthesized using reverse transcriptase and the RT primers described above (Unisplint-S or flagASA), following the manufacturer's instructions.

The PROfusion™ obtained for each fragment was next purified on the resin appropriate to its peptide purification tag, i.e., on Ni-NTA agarose for 10 the His₆-tag and M2 agarose for the FLAG-tag, following the procedure recommended by the manufacturer. The DNA component of the tag-binding PROfusions™ was amplified by PCR using Pharmacia Ready-to-Go PCR Beads, 10 pmol of 5' and 3' PCR primers, and the following PCR program (Pharmacia, Piscataway, NJ): Step 1: 95°C for 3 minutes; Step 2: 95°C for 30 15 seconds, 58/62°C for 30 seconds, 72°C for 1 minute, 20/25/30 cycles, as required; Step 3: 72°C for 5 minutes; Step 4: 4°C until end.

The resulting DNA was cleaved by 5 U EarI (New England Biolabs) per 1 ug DNA; the reaction took place in T4 DNA Ligase Buffer (New England Biolabs) at 37°C, for 1 hour, and was followed by an incubation at 70°C for 15 20 minutes to inactivate Ear I. Equal amounts of the BC, DE, and FG fragments were combined and ligated to form a full-length ¹⁰F_n3 gene with randomized loops. The ligation required 10 U of fresh EarI (New England Biolabs) and 20 U of T4 DNA Ligase (Promega, Madison, WI), and took 1 hour at 37°C.

Three different libraries were made in the manner described above. 25 Each contained the form of the FG loop with 10 randomized residues. The BC and the DE loops of the first library bore the wild type ¹⁰F_n3 sequence; a BC loop with 7 randomized residues and a wild type DE loop made up the second

library; and a BC loop with 7 randomized residues and a DE loop with 4 randomized residues made up the third library. The complexity of the FG loop in each of these three libraries was 10^{13} ; the further two randomized loops provided the potential for a complexity too large to be sampled in a laboratory.

5 The three libraries constructed were combined into one master library in order to simplify the selection process; target binding itself was expected to select the most suitable library for a particular challenge. PROfusionsTM were obtained from the master library following the general procedure described in Szostak et al., U.S.S.N. 09/007,005 and 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302 (Figure 8).

Fusion Selections

15 The master library in the PROfusionTM form was subjected to selection for binding to TNF- α . Two protocols were employed: one in which the target was immobilized on an agarose column and one in which the target was immobilized on a BIACORE chip. First, an extensive optimization of conditions to minimize background binders to the agarose column yielded the favorable buffer conditions of 50 mM HEPES pH 7.4, 0.02% Triton, 100 μ g/ml Sheared Salmon Sperm DNA. In this buffer, the non-specific binding of the ¹⁰F_n3 RNA fusion to TNF- α Sepharose was 0.3%. The non-specific binding background of the ¹⁰F_n3 RNA-DNA to TNF- α Sepharose was found to be 0.1%.

25 During each round of selection on TNF- α Sepharose, the ProfusionTM library was first preincubated for an hour with underivatized Sepharose to remove any remaining non-specific binders; the flow-through from this pre-clearing was incubated for another hour with TNF- α Sepharose.

The TNF- α Sepharose was washed for 3-30 minutes.

After each selection, the PROfusion™ DNA that had been eluted from the solid support with 0.3 M NaOH or 0.1M KOH was amplified by PCR; a DNA band of the expected size persisted through multiple rounds of selection (Figure 9); similar results were observed in the two alternative selection protocols, and only the data from the agarose column selection is shown in Figure 9.

In the first seven rounds, the binding of library PROfusions™ to the target remained low; in contrast, when free protein was translated from DNA pools at different stages of the selection, the proportion of the column binding species increased significantly between rounds (Figure 10). Similar selections may be carried out with any other binding species target (for example, IL-1 and IL-13).

Animal Studies

Wild-type ¹⁰F_n3 contains an integrin-binding tripeptide motif, Arginine 78 - Glycine 79 - Aspartate 80 (the "RGD motif") at the tip of the FG loop. In order to avoid integrin binding and a potential inflammatory response based on this tripeptide in vivo, a mutant form of ¹⁰F_n3 was generated that contained an inert sequence, Serine 78 - Glycine 79 - Glutamate 80 (the "SGE mutant"), a sequence which is found in the closely related, wild-type ¹¹F_n3 domain. This SGE mutant was expressed as an N-terminally His₆-tagged, free protein in E. coli, and purified to homogeneity on a metal chelate column followed by a size exclusion column.

In particular, the DNA sequence encoding His₆-¹⁰F_n3(SGE) was cloned into the pET9a expression vector and transformed into BL21 DE3 pLysS cells. The culture was then grown in LB broth containing 50 μ g/mL

kanamycin at 37°C, with shaking, to $A_{560}=1.0$, and was then induced with 0.4 mM IPTG. The induced culture was further incubated, under the same conditions, overnight (14-18 hours); the bacteria were recovered by standard, low speed centrifugation. The cell pellet was resuspended in 1/50 of the original culture volume of lysis buffer (50 mM Tris 8.0, 0.5 M NaCl, 5% glycerol, 0.05% Triton X-100, and 1 mM PMSF), and the cells were lysed by passing the resulting paste through a Microfluidics Corporation Microfluidizer M110-EH, three times. The lysate was clarified by centrifugation, and the supernatant was filtered through a 0.45 μ m filter followed by filtration through a 0.2 μ m filter. 100 mL of the clarified lysate was loaded onto a 5 mL Talon cobalt column (Clontech, Palo Alto, CA), washed by 70 mL of lysis buffer, and eluted with a linear gradient of 0-30 mM imidazole in lysis buffer. The flow rate through the column through all the steps was 1 mL/min. The eluted protein was concentrated 10-fold by dialysis (MW cutoff = 3,500) against 15,000-20,000 PEG. The resulting sample was dialysed into buffer 1 (lysis buffer without the glycerol), then loaded, 5 mL at a time, onto a 16 x 60 mm Sephacryl 100 size exclusion column equilibrated in buffer 1. The column was run at 0.8 mL/min, in buffer 1; all fractions that contained a protein of the expected MW were pooled, concentrated 10X as described above, then dialyzed into PBS. Toxikon (MA) was engaged to perform endotoxin screens and animal studies on the resulting sample.

In these animal studies, the endotoxin levels in the samples examined to date have been below the detection level of the assay. In a preliminary toxicology study, this protein was injected into two mice at the estimated 100X therapeutic dose of 2.6 mg/mouse. The animals survived the two weeks of the study with no apparent ill effects. These results suggest that $^{10}\text{Fn3}$ may be incorporated safely into an IV drug.

Alternative Constructs for In Vivo Use

To extend the half life of the 8 kD $^{10}\text{Fn3}$ domain, a larger molecule has also been constructed that mimics natural antibodies. This $^{10}\text{Fn3-Fc}$ molecule contains the $-\text{CH}_1-\text{CH}_2-\text{CH}_3$ (Figure 11) or $-\text{CH}_2-\text{CH}_3$ domains of the IgG constant region of the host; in these constructs, the $^{10}\text{Fn3}$ domain is grafted onto the N-terminus in place of the IgG V_H domain (Figures 11 and 12). Such antibody-like constructs are expected to improve the pharmacokinetics of the protein as well as its ability to harness the natural immune response.

In order to construct the murine form of the $^{10}\text{Fn3-CH}_1-\text{CH}_2-\text{CH}_3$ clone, the $-\text{CH}_1-\text{CH}_2-\text{CH}_3$ region was first amplified from a mouse liver spleen cDNA library (Clontech), then ligated into the pET25b vector. The primers used in the cloning were 5' Fc Nest and 3' 5 Fc Nest, and the primers used to graft the appropriate restriction sites onto the ends of the recovered insert were 5' Fc HIII and 3' Fc Nhe:

5' Fc Nest 5'GCG GCA GGG TTT GCT TAC TGG GGC CAA GGG 3' (SEQ ID NO: 15);
 3' Fc Nest 5'GGG AGG GGT GGA GGT AGG TCA CAG TCC 3' (SEQ ID NO: 16);
 3' Fc Nhe 5' TTT GCT AGC TTT ACC AGG AGA GTG GGA GGC 3' (SEQ ID NO: 17); and
 5' Fc HIII 5' AAA AAG CTT GCC AAA ACG ACA CCC CCA TCT GTC 3' (SEQ ID NO: 18).

Further PCR is used to remove the CH_1 region from this clone and create the Fc part of the shorter, $^{10}\text{Fn3-CH}_2-\text{CH}_3$ clone. The sequence encoding $^{10}\text{Fn3}$ is spliced onto the 5' end of each clone; either the wild type

¹⁰F_n3 cloned from the same mouse spleen cDNA library or a modified ¹⁰F_n3 obtained by mutagenesis or randomization of the molecules can be used. The oligonucleotides used in the cloning of murine wild-type ¹⁰F_n3 were:

Mo 5PCR-NdeI:

5 5' CATATGGTTTCTGATATTCCGAGAGATCTGGAG 3' (SEQ ID NO: 19);

Mo5PCR-His-NdeI (for an alternative N-terminus with the His₆ purification tag):

5' CAT ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT
10 ATT CCG AGA G 3' (SEQ ID NO: 20); and

Mo3PCR-EcoRI: 5'

GAATTCCTATGTTTTATAATTGATGGAAAC3' (SEQ ID NO: 21).

The human equivalents of the clones are constructed using the same strategy with human oligonucleotide sequences.

15 ¹⁰F_n3 Scaffolds in Protein Chip Applications

The suitability of the ¹⁰F_n3 scaffold for protein chip applications is the consequence of (1) its ability to support many binding functions which can be selected rapidly on the bench or in an automated setup, and (2) its superior biophysical properties.

20 The versatile binding properties of ¹⁰F_n3 are a function of the loops displayed by the F_n3 immunoglobulin-like, beta sandwich fold. As discussed above, these loops are similar to the complementarity determining regions of antibody variable domains and can cooperate in a way similar to those antibody loops in order to bind antigens. In our system, ¹⁰F_n3 loops BC (residues

21-30), DE (residues 51-56), and FG (residues 76-87) are randomized either in sequence, in length, or in both sequence and length in order to generate diverse libraries of mRNA-¹⁰F_n3 fusions. The binders in such libraries are then enriched based on their affinity for an immobilized or tagged target, until a
5 small population of high affinity binders are generated. Also, error-prone PCR and recombination can be employed to facilitate affinity maturation of selected binders. Due to the rapid and efficient selection and affinity maturation protocols, binders to a large number of targets can be selected in a short time.

As a scaffold for binders to be immobilized on protein chips, the
10 ¹⁰F_n3 domain has the advantage over antibody fragments and single-chain antibodies of being smaller and easier to handle. For example, unlike single-chain scaffolds or isolated variable domains of antibodies, which vary widely in their stability and solubility, and which require an oxidizing environment to preserve their structurally essential disulfide bonds, ¹⁰F_n3 is
15 extremely stable, with a melting temperature of 110°C, and solubility at a concentration > 16 mg/mL. The ¹⁰F_n3 scaffold also contains no disulfides or free cysteines; consequently, it is insensitive to the redox potential of its environment. A further advantage of ¹⁰F_n3 is that its antigen-binding loops and N-terminus are on the edge of the beta-sandwich opposite to the
20 C-terminus; thus the attachment of a ¹⁰F_n3 scaffold to a chip by its C-terminus aligns the antigen-binding loops, allowing for their greatest accessibility to the solution being assayed. Since ¹⁰F_n3 is a single domain of only 94 amino acid residues, it is also possible to immobilize it onto a chip surface at a higher
25 density than is used for single-chain antibodies, with their approximately 250 residues. In addition, the hydrophilicity of the ¹⁰F_n3 scaffold, which is reflected in the high solubility of this domain, leads to a lower than average background binding of ¹⁰F_n3 to a chip surface.

The stability of the $^{10}\text{Fn3}$ scaffold as well as its suitability for library formation and selection of binders are likely to be shared by the large, Fn3-like class of protein domains with an immunoglobulin-like fold, such as the domains of tenascin, N-cadherin, E-cadherin, ICAM, titin, GCSF-R, cytokine receptor, glycosidase inhibitor, and antibiotic chromoprotein. The key features shared by all such domains are a stable framework provided by two beta-sheets, which are packed against each other and which are connected by at least three solvent-accessible loops per edge of the sheet; such loops can be randomized to generate a library of potential binders without disrupting the structure of the framework (as described above).

Immobilization of Fibronectin Scaffold Binders (Fn-binders)

To immobilize Fn-binders to a chip surface, a number of exemplary techniques may be utilized. For example, Fn-binders may be immobilized as RNA-protein fusions by Watson-Crick hybridization of the RNA moiety of the fusion to a base complementary DNA immobilized on the chip surface (as described, for example, in Addressable Protein Arrays, U.S.S.N. 60/080,686; U.S.S.N. 09/282,734; and WO 99/51773). Alternatively, Fn-binders can be immobilized as free proteins directly on a chip surface. Manual as well as robotic devices may be used for deposition of the Fn-binders on the chip surface. Spotting robots can be used for deposition of Fn-binders with high density in an array format (for example, by the method of Lueking et al., Anal Biochem. 1999 May 15;270(1):103-11). Different methods may also be utilized for anchoring the Fn-binder on the chip surface. A number of standard immobilization procedures may be used including those described in Methods in Enzymology (K. Mosbach and B. Danielsson, eds.), vols. 135 and 136, Academic Press, Orlando, Florida, 1987; Nilsson et al., Protein Expr. Purif.

1997 Oct;11(1):1-16; and references therein. Oriented immobilization of Fn-binders can help to increase the binding capacity of chip-bound Fn-binders. Exemplary approaches for achieving oriented coupling are described in Lu et al., The Analyst (1996), vol. 121, p. 29R-32R; and Turkova, J Chromatogr B Biomed Sci App. 1999 Feb 5;722(1-2):11-31. In addition, any of the methods described herein for anchoring Fn-binders to chip surfaces can also be applied to the immobilization of Fn-binders on beads, or other supports.

Target Protein Capture and Detection

Selected populations of Fn-binders may be used for detection and/or quantitation of analyte targets, for example, in samples such as biological samples. To carry out this type of diagnostic assay, selected Fn-binders to targets of interest are immobilized on an appropriate support to form multi-featured protein chips. Next, a sample is applied to the chip, and the components of the sample that associate with the Fn-binders are identified based on the target-specificity of the immobilized binders. Using this technique, one or more components may be simultaneously identified or quantitated in a sample (for example, as a means to carry out sample profiling).

Methods for target detection allow measuring the levels of bound protein targets and include, without limitation, radiography, fluorescence scanning, mass spectroscopy (MS), and surface plasmon resonance (SPR). Autoradiography using a phosphorimager system (Molecular Dynamics, Sunnyvale, CA) can be used for detection and quantification of target protein which has been radioactively labeled, e.g., using ³⁵S methionine. Fluorescence scanning using a laser scanner (see below) may be used for detection and quantification of fluorescently labeled targets. Alternatively, fluorescence scanning may be used for the detection of fluorescently labeled ligands which

themselves bind to the target protein (e.g., fluorescently labeled target-specific antibodies or fluorescently labeled streptavidin binding to target-biotin, as described below).

Mass spectroscopy can be used to detect and identify bound targets based on their molecular mass. Desorption of bound target protein can be achieved with laser assistance directly from the chip surface as described below. Mass detection also allows determinations, based on molecular mass, of target modifications including post-translational modifications like phosphorylation or glycosylation. Surface plasmon resonance can be used for quantification of bound protein targets where the Fn-binder(s) are immobilized on a suitable gold-surface (for example, as obtained from Biacore, Sweden).

Described below are exemplary schemes for selecting Fn binders (in this case, Fn-binders specific for the protein, TNF- α) and the use of those selected populations for detection on chips. This example is provided for the purpose of illustrating the invention, and should not be construed as limiting.

Selection of TNF- α Binders Based on ¹⁰Fn3 Scaffold

In one exemplary use for fibronectin scaffold selection on chips, an ¹⁰Fn3-based selection was performed against TNF- α , using a library of human ¹⁰Fn3 variants with randomized loops BC, DE, and FG. The library was constructed from three DNA fragments, each of which contained nucleotide sequences that encoded approximately one third of human ¹⁰Fn3, including one of the randomized loops. The DNA sequences that encoded the loop residues listed above were rebuilt by oligonucleotide synthesis, so that the codons for the residues of interest were replaced by (NNS)_n, where N represents any of the four deoxyribonucleotides (A, C, G, or T), and S represents either C or G. The C-terminus of each fragment contained the sequence for the FLAG

purification tag.

Once extended by Klenow, each DNA fragment was transcribed, ligated to a puromycin-containing DNA linker, and translated in vitro, as described by Szostak et al. (Roberts and Szostak, Proc. Natl. Acad. Sci USA 94:12297, 1997; Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700), to generate an mRNA-peptide fusion, which was then reverse-transcribed into a DNA-mRNA-peptide fusion. The binding of the FLAG-tagged peptide to M2 agarose separated full-length fusion molecules from those containing frameshifts or superfluous stop codons; the DNA associated with the purified full-length fusion was amplified by PCR, then the three DNA fragments were cut by Ear I restriction endonuclease and ligated to form the full length template. The template was transcribed, ligated to puromycin-containing DNA linkers, and translated to generate a ¹⁰F_n3-PROfusion™ library, which was then reverse-transcribed to yield the DNA-mRNA-peptide fusion library which was subsequently used in the selection.

Selection for TNF- α binders took place in 50 mM HEPES, pH 7.4, 0.02% Triton-X, 0.1 mg/mL salmon sperm DNA. The PROfusion™ library was incubated with Sepharose-immobilized TNF- α ; after washing, the DNA associated with the tightest binders was eluted with 0.1 M KOH, amplified by PCR, and transcribed, ligated, translated, and reverse-transcribed into the starting material for the next round of selection.

Ten rounds of such selection were performed (as shown in Figure 13); they resulted in a PROfusion™ pool that bound to TNF- α -Sepharose with the apparent average K_d of 120 nM. Specific clonal components of the pool that were characterized showed TNF- α binding in the range of 50-500 nM.

Fn-binder Immobilization, Target Protein Capture, and MALDI-TOF

Detection

As a first step toward immobilizing the Fn-binders to a chip surface, an oligonucleotide capture probe was prepared with an automated DNA synthesizer (PE BioSystems Expedite 8909) using the solid-support phosphoramidite approach. All reagents were obtained from Glen Research. Synthesis was initiated with a solid support containing a disulfide bond to eventually provide a 3'-terminal thiol functionality. The first four monomers to be added were hexaethylene oxide units, followed by 20 T monomers. The 5'-terminal DMT group was not removed. The capture probe was cleaved from the solid support and deprotected with ammonium hydroxide, concentrated to dryness in a vacuum centrifuge, and purified by reverse-phase HPLC using an acetonitrile gradient in triethylammonium acetate buffer. Appropriate fractions from the HPLC were collected, evaporated to dryness in a vacuum centrifuge, and the 5'-terminal DMT group was removed by treatment with 80% AcOH for 30 minutes. The acid was removed by evaporation, and the oligonucleotide was then treated with 100 mM DTT for 30 minutes to cleave the disulfide bond. DTT was removed by repeated extraction with EtOAc. The oligonucleotide was ethanol precipitated from the remaining aqueous layer and checked for purity by reverse-phase HPLC.

The 3'-thiol capture probe was adjusted to 250 μ M in degassed 1X PBS buffer and applied as a single droplet (75 μ L) to a 9x9mm gold-coated chip (Biacore) in an argon-flushed chamber containing a small amount of water. After 18 hours at room temperature, the capture probe solution was removed, and the functionalized chip was washed with 50 mL 1X PBS buffer (2x for 15 minutes each) with gentle agitation, and then rinsed with 50 mL water (2x for 15 minutes each) in the same fashion. Remaining liquid was

carefully removed and the functionalized chips were either used immediately or stored at 4°C under argon.

About 1pmol of ¹⁰F_n3 fusion pool from the Round 10 TNF-α selection (above) was treated with RNase A for several hours, adjusted to 5X SSC in 70 μL, and applied to a functionalized gold chip from above as a single droplet. A 50 μL volume gasket device was used to seal the fusion mixture with the functionalized chip, and the apparatus was continuously rotated at 4°C. After 18 hours the apparatus was disassembled, and the gold chip was washed with 50 mL 5X SSC for 10 minutes with gentle agitation. Excess liquid was carefully removed from the chip surface, and the chip was passivated with a blocking solution (1X TBS + 0.02% Tween-20 + 0.25% BSA) for 10 minutes at 4°C. Excess liquid was carefully removed, and a solution containing 500 μg/mL TNF-α in the same composition blocking solution was applied to the chip as a single droplet and incubated at 4°C for two hours with occasional mixing of the droplet via Pipetman. After removal of the binding solution, the chip was washed for 5 minutes at 4°C with gentle agitation (50 mL 1X TBS + 0.02% Tween-20) and then dried at room temperature. A second chip was prepared exactly as described above, except fusion was not added to the hybridization mix.

Next, MALDI-TOF matrix (15 mg/mL 3,5-dimethoxy-4-hydroxycinnamic acid in 1:1 ethanol/10% formic acid in water) was uniformly applied to the gold chips with a high-precision 3-axis robot (MicroGrid, BioRobotics). A 16-pin tool was used to transfer the matrix from a 384-well microtiter plate to the chips, producing 200 micron diameter features with a 600 micron pitch. The MALDI-TOF mass spectrometer (Voyager DE, PerSeptive Biosystems) instrument settings were as follows: Accelerating Voltage = 25k, Grid Voltage = 92%, Guide Wire Voltage =

0.05%, Delay = 200 on, Laser Power = 2400, Low Mass Gate = 1500, Negative Ions = off. The gold chips were individually placed on a MALDI sample stage modified to keep the level of the chip the same as the level of the stage, thus allowing proper flight distance. The instrument's video monitor and motion control system were used to direct the laser beam to individual matrix features.

Figures 14 and 15 show the mass spectra from the $^{10}\text{Fn3}$ fusion chip and the non-fusion chip, respectively. In each case, a small number of 200 micron features were analyzed to collect the spectra, but Figure 15 required significantly more acquisitions. The signal at 17.5 kDa corresponds to TNF- α monomer.

Fn-binder Immobilization, Target Protein Capture, and Fluorescence Detection

Pre-cleaned 1x3 inch glass microscope slides (Goldseal, #3010) were treated with Nanostrip (Cyantek) for 15 minutes, 10% aqueous NaOH at 70°C for 3 minutes, and 1% aqueous HCl for 1 minute, thoroughly rinsing with deionized water after each reagent. The slides were then dried in a vacuum desiccator over anhydrous calcium sulfate for several hours. A 1% solution of aminopropyltrimethoxysilane in 95% acetone / 5% water was prepared and allowed to hydrolyze for 20 minutes. The glass slides were immersed in the hydrolyzed silane solution for 5 minutes with gentle agitation. Excess silane was removed by subjecting the slides to ten 5-minute washes, using fresh portions of 95% acetone / 5% water for each wash, with gentle agitation. The slides were then cured by heating at 110°C for 20 minutes. The silane treated slides were immersed in a freshly prepared 0.2% solution of phenylene 1,4-diisothiocyanate in 90% DMF / 10% pyridine for two hours, with gentle agitation. The slides were washed sequentially with 90% DMF /

10% pyridine, methanol, and acetone. After air drying, the functionalized slides were stored at 0°C in a vacuum desiccator over anhydrous calcium sulfate. Similar results were obtained with commercial amine-reactive slides (3-D Link, Surmodics).

5 Oligonucleotide capture probes were prepared with an automated DNA synthesizer (PE BioSystems Expedite 8909) using conventional phosphoramidite chemistry. All reagents were from Glen Research. Synthesis was initiated with a solid support bearing an orthogonally protected amino functionality, whereby the 3'-terminal amine is not unmasked until final
10 deprotection step. The first four monomers to be added were hexaethylene oxide units, followed by the standard A, G, C and T monomers. All capture oligo sequences were cleaved from the solid support and deprotected with ammonium hydroxide, concentrated to dryness, precipitated in ethanol, and purified by reverse-phase HPLC using an acetonitrile gradient in
15 triethylammonium acetate buffer. Appropriate fractions from the HPLC were collected, evaporated to dryness in a vacuum centrifuge, and then coevaporated with a portion of water.

 The purified, amine-labeled capture oligos were adjusted to a concentration of 250 μ M in 50 mM sodium carbonate buffer (pH 9.0)
20 containing 10% glycerol. The probes were spotted onto the amine-reactive glass surface at defined positions in a 5x5x6 array pattern with a 3-axis robot (MicroGrid, BioRobotics). A 16-pin tool was used to transfer the liquid from 384-well microtiter plates, producing 200 micron features with a 600 micron pitch. Each sub-grid of 24 features represents a single capture probe (i.e., 24
25 duplicate spots). The arrays were incubated at room temperature in a moisture-saturated environment for 12-18 hours. The attachment reaction was terminated by immersing the chips in 2% aqueous ammonium hydroxide for

five minutes with gentle agitation, followed by rinsing with distilled water (3X for 5 minutes each). The array was finally soaked in 10X PBS solution for 30 minutes at room temperature, and then rinsed again for 5 minutes in distilled water.

5 Specific and thermodynamically isoenergetic sequences along the ¹⁰F_n3 mRNA were identified to serve as capture points to self-assemble and anchor the ¹⁰F_n3 protein. The software program HybSimulator v4.0 (Advanced Gene Computing Technology, Inc.) facilitated the identification and analysis of potential capture probes. Six unique capture probes were
10 chosen and printed onto the chip, three of which are complementary to common regions of the ¹⁰F_n3 fusion pool's mRNA (CP3', CP5', and CPflag). The remaining three sequences (CPneg1, CPneg2, and CPneg3) are not complementary and function in part as negative controls. Each of the capture probes possesses a 3'-amino terminus and four hexaethylene oxide spacer units,
15 as described above. The following is a list of the capture probe sequences that were employed (5'→3'):

CP3': TGTAATAGTAATTGTCCC (SEQ ID NO: 22)
CP5': TTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 23)
CPneg1: CCTGTAGGTGTCCAT (SEQ ID NO: 24)
20 CPflag: CATCGTCCTTGTAGTC (SEQ ID NO: 25)
CPneg2: CGTCGTAGGGGTA (SEQ ID NO: 26)
CPneg3: CAGGTCTTCTTCAGAGA (SEQ ID NO: 27)

About 1pmol of ¹⁰F_n3 fusion pool from the Round 10 TNF-α selection was adjusted to 5X SSC containing 0.02% Tween-20 and 2 mM vanadyl
25 ribonucleotide complex in a total volume of 350 μL. The entire volume was

applied to the microarray under a 400 μ L gasket device and the assembly was continuously rotated for 18 hours at room temperature. After hybridization the slide was washed sequentially with stirred 500 mL portions of 5X SSC, 2.5X SSC, and 1X SSC for 5 minutes each. Traces of liquid were removed by
5 centrifugation and the slide was allowed to air-dry.

Recombinant human TNF- α (500 μ g, lyophilized, from PreproTech) was taken up in 230 μ L 1X PBS and dialyzed against 700 mL stirred 1X PBS at 4°C for 18 hours in a Microdialyzer unit (3,500 MWCO, Pierce). The dialyzed TNF- α was treated with EZ-Link NHS-LC-LC biotinylation reagent
10 (20 μ g, Pierce) for 2 hours at 0°C, and again dialyzed against 700 mL stirred 1X PBS at 4°C for 18 hours in a Microdialyzer unit (3,500 MWCO, Pierce). The resulting conjugate was analyzed by MALDI-TOF mass spectrometry and was found to be almost completely functionalized with a single biotin moiety.

Each of the following processes was conducted at 4°C with
15 continuous rotation or mixing. The protein microarray surface was passivated by treatment with 1X TBS containing 0.02% Tween-20 and 0.2% BSA (200 μ L) for 60 minutes. Biotinylated TNF- α (100 nM concentration made up in the passivation buffer) was contacted with the microarray for 120 minutes. The microarray was washed with 1X TBS containing 0.02% Tween-20 (3X 50
20 mL, 5 minutes each wash). Fluorescently labeled streptavidin (2.5 μ g/mL Alexa 546-streptavidin conjugate from Molecular Probes, made up in the passivation buffer) was contacted with the microarray for 60 minutes. The microarray was washed with 1X TBS containing 0.02% Tween-20 (2X 50 mL, 5 minutes each wash) followed by a 3 minute rinse with 1X TBS. Traces of
25 liquid were removed by centrifugation, and the slide was allowed to air-dry at room temperature.

Fluorescence laser scanning was performed with a GSI Lumonics ScanArray 5000 system using 10 μ M pixel resolution and preset excitation and emission wavelengths for Alexa 546 dye. Phosphorimage analysis was performed with a Molecular Dynamics Storm system. Exposure time was 48
5 hours with direct contact between the microarray and the phosphor storage screen. Phosphorimage scanning was performed at the 50 μ M resolution setting, and data was extracted with ImageQuant v.4.3 software.

Figures 16 and 17 are the phosphorimage and fluorescence scan, respectively, of the same array. The phosphorimage shows where the $^{10}\text{Fn3}$
10 fusion hybridized based on the ^{35}S methionine signal. The fluorescence scan shows where the labeled TNF- α bound.

Other Embodiments

Other embodiments are within the claims.

All publications, patents, and patent applications mentioned herein
15 are hereby incorporated by reference.

What is claimed is:

Claims

1. An array of proteins immobilized on a solid support, each of said proteins comprising a fibronectin type III domain having at least one randomized loop, at least one randomized β -sheet, or a combination thereof,
5 and being characterized by its ability to bind to a compound that is not bound by a corresponding naturally-occurring fibronectin.

2. The array of claim 1, wherein said fibronectin type III domain is a mammalian fibronectin type III domain.

3. The array of claim 2, wherein said fibronectin type III domain is a
10 human fibronectin type III domain.

4. The array of claim 1, wherein each of said proteins comprises the tenth module of said fibronectin type III domain ($^{10}\text{Fn3}$).

5. The array of claim 4, wherein each of said proteins contains one, two, or three randomized loops and wherein at least one of said loops
15 contributes to the binding of the protein to said compound.

6. The array of claim 5, wherein at least two of said randomized loops contribute to said binding of the protein to said compound.

7. The array of claim 6, wherein at least three of said randomized loops contribute to said binding of the protein to said compound.

8. The array of claim 4, wherein said ¹⁰F_n3 lacks an integrin-binding motif.

9. The array of claim 1, wherein each of said proteins lacks disulfide bonds.

5 10. The array of claim 1, wherein each of said proteins is a monomer or a dimer.

11. The array of claim 1, wherein each of said proteins is covalently bound to a nucleic acid.

10 12. The array of claim 11, wherein said nucleic acid encodes the covalently bound protein.

13. The array of claim 12, wherein said nucleic acid is RNA.

14. The array of claim 1, wherein said solid support is a chip.

15. A method for obtaining a protein which binds to a compound, said method comprising:

15 (a) contacting said compound with an array of candidate proteins immobilized on a solid support, each of said candidate proteins comprising a fibronectin type III domain having at least one randomized loop, one randomized β-sheet, or a combination thereof, said contacting being carried out under conditions that allow compound-protein complex formation; and
20 (b) obtaining, from said complex, a protein which binds to said compound.

16. A method for obtaining a compound which binds to a protein, said protein comprising a fibronectin type III domain having at least one randomized loop, at least one randomized β -sheet, or a combination thereof, said method comprising:

5 (a) contacting an array of proteins immobilized on a solid support with a candidate compound, each of said proteins comprising a fibronectin type III domain having at least one randomized loop, one randomized β -sheet, or a combination thereof, said contacting being carried out under conditions that allow compound-protein complex formation; and

10 (b) obtaining, from said complex, a compound which binds to a protein of the array.

17. The method of claim 15, said method further comprising the steps of:

15 (c) further randomizing a protein which binds to said compound in step (b);

(d) forming an array on a solid support with the further randomized proteins of step (c); and

(e) repeating steps (a) and (b) using, in step (a), the array of further randomized proteins as said array of candidate proteins.

20 18. The method of claim 16, said method further comprising the steps of:

(c) modifying the compound which binds to said protein in step (b);

and

25 (d) repeating steps (a) and (b) using, in step (a), said further modified compound as said candidate compound.

19. The method of claim 15 or 16, wherein said solid support is a chip.

20. A method for detecting a compound in a sample, said method comprising:

5 (a) contacting a sample with a protein which binds to said compound and which comprises a fibronectin type III domain having at least one randomized loop, at least one randomized β -sheet, or a combination thereof, said contacting being carried out under conditions that allow compound-protein complex formation; and

10 (b) detecting said complex, thereby detecting said compound in said sample.

21. The method of claim 20, wherein said sample is a biological sample.

22. The method of claim 20, wherein said protein is immobilized on
15 a solid support.

23. The method of claim 22, wherein said protein is immobilized on said solid support as part of an array.

24. The method of claim 22, wherein said solid support is a bead or chip.

20 25. The method of claim 15, 16 or 20, wherein said compound is a protein.

26. The method of claim 15, 16, or 20, wherein said fibronectin type III domain is a mammalian fibronectin type III domain.

27. The method of claim 26, wherein said fibronectin type III domain is a human fibronectin type III domain.

5 28. The method of claim 15, 16, or 20, wherein each of said proteins comprises the tenth module of said fibronectin type III domain (¹⁰Fn3).

29. The method of claim 28, wherein each of said proteins contains one, two, or three, randomized loops and wherein at least one of said loops contributes to the binding of said protein to said compound.

10 30. The method of claim 28, wherein said ¹⁰Fn3 lacks an integrin-binding motif.

31. The method of claim 15, 16, or 20, wherein each of said proteins is covalently bound to a nucleic acid.

15 32. The method of claim 31, wherein said nucleic acid encodes the covalently bound protein.

33. The method of claim 32, wherein said nucleic acid is RNA.

34. The method of claim 15, 16, or 20, wherein said complex or said compound is detected by radiography, fluorescence detection, mass spectroscopy, or surface plasmon resonance.

1/15

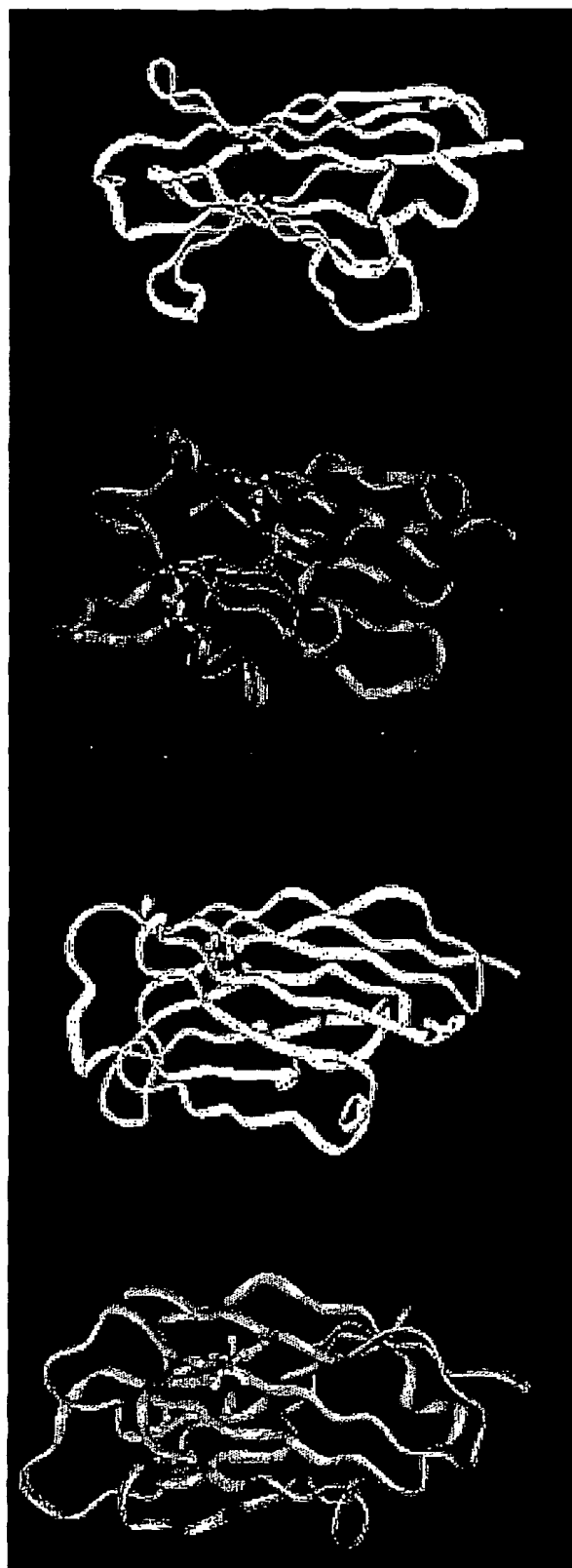


Fig. 1

2/15

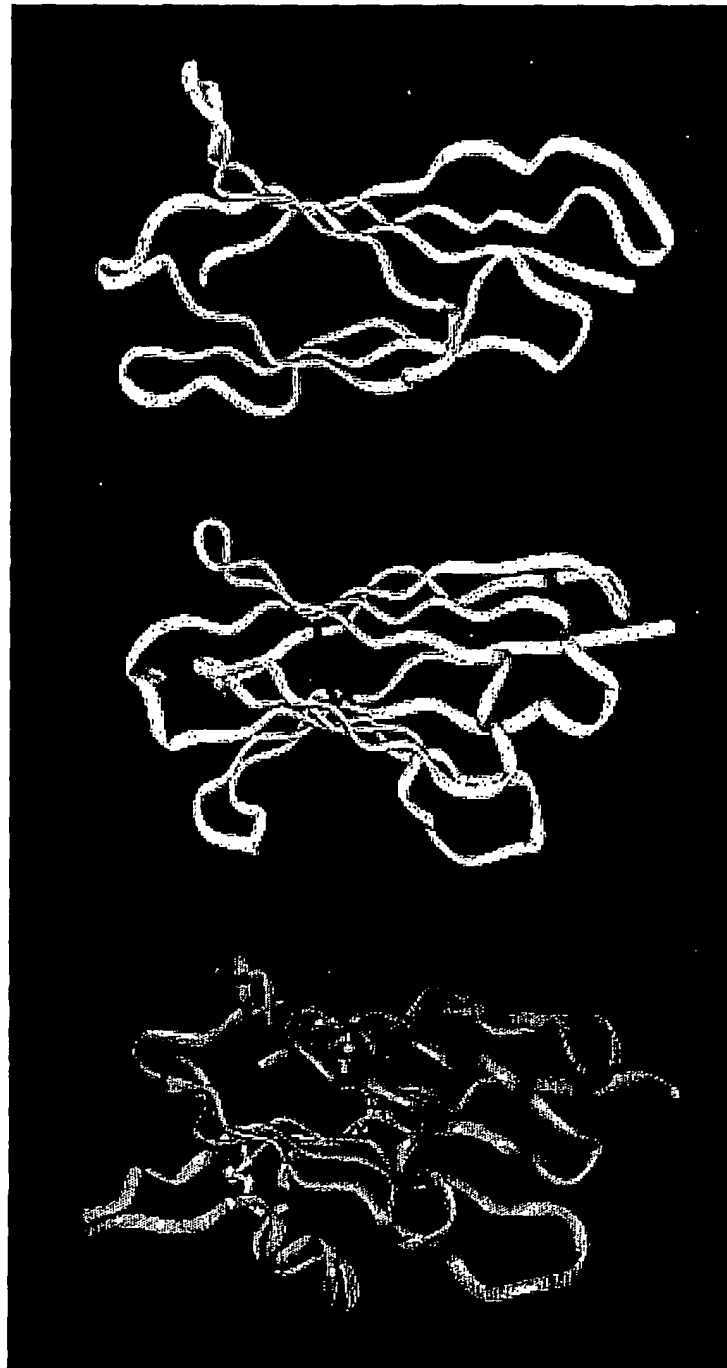


Fig. 2

3/15

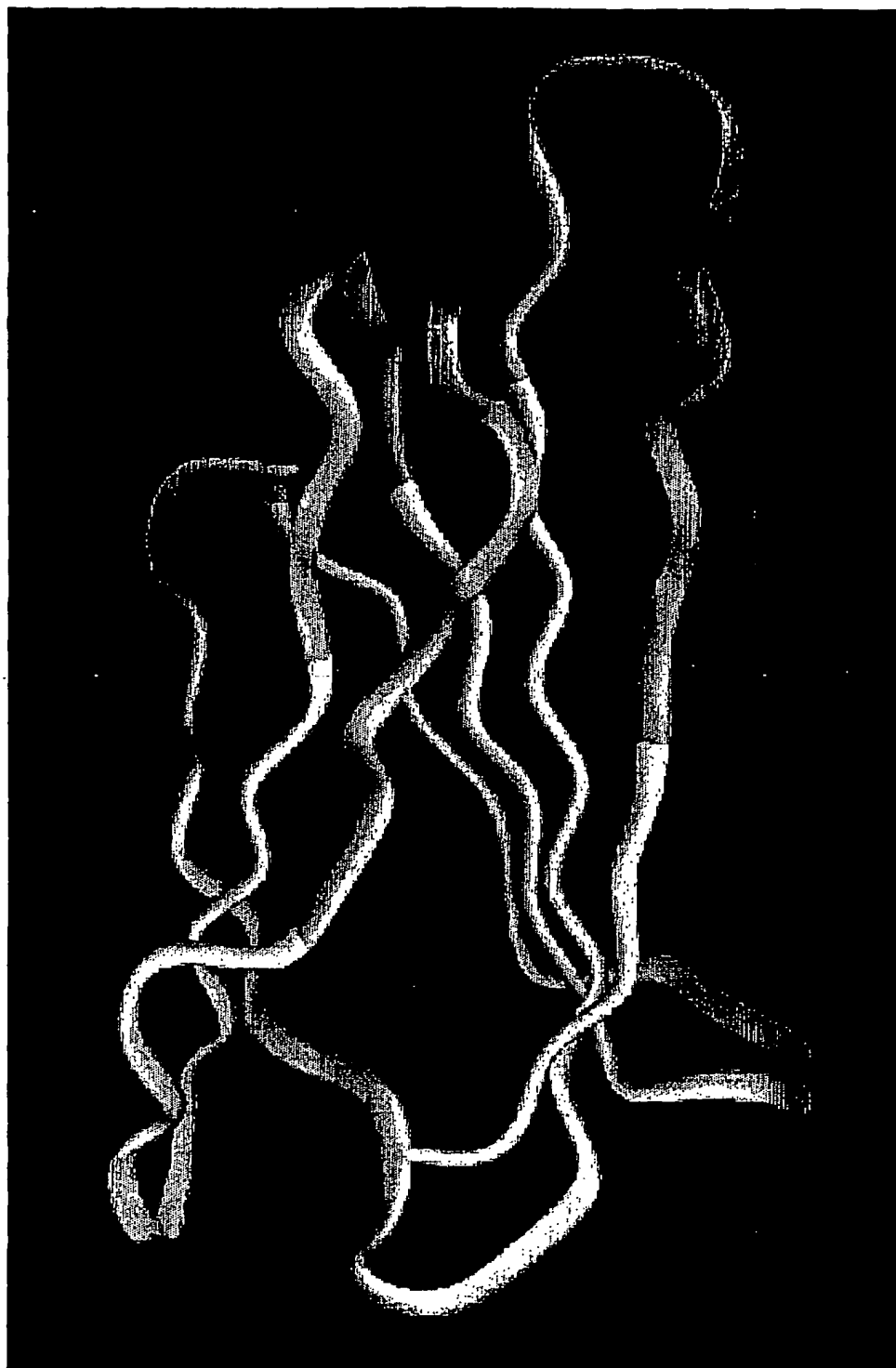


Fig. 3

4/15

Hs FND	1	9	10	19	20	29	30	37	38	47	48	57	58	67	68	77	78	87	88	94
Bt FN	VSDVPRD-LE	VVAATPTSL	ISWDAPAVTV	RYRITYG--	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ
Rn FN	VSDVPRD-LE	VVAATPTSL	ISWDAPAVTV	RYRITYG--	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ
Mm FN	VSDVPRD-LE	VVAATPTSL	ISWDAPAVTV	RYRITYG--	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ
Oc FN	VSDVPRD-LE	VVAATPTSL	ISWDAPAVTV	RYRITYG--	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ
Gg FN	VSDVPRD-LE	VVAATPTSL	ISWDAPAVTV	RYRITYG--	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ
Xl FN	VSDVPRD-LE	VVAATPTSL	ISWDAPAVTV	RYRITYG--	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ
Cf FN	AIDAPSn-Lr	FLATTPnSL	ISWDAPAVTV	RYRITYG--	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ
Ec FN	AIDAPSn-Lh	FLATTPnSL	ISWDAPAVTV	RYRITYG--	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ
Hs TC	VS-PPKD-Lv	VTeVTeTVN	LAWDn-eMrV	TeVILVYTP	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM
Ss TP	VS-PPKD-Li	VTeVTeTVN	LAWDn-eMrV	TeVILVYTP	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM	THEGGLM
Mm TX	MIDGPQD-Lr	VVAATPTSL	ISWDAPAVTV	RYRITYG--	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ	ETGNSPVQ
Hs CAP	TlpVPvVSLn	IYdVGPtTMH	VQWOp-VGGA	TGYILSYkPV	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE
Oc C12	TlpVPvVSLn	IYdVGPtTMH	VQWOp-VGGA	TGYILSYkPV	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE	KDTpTrpKE
Gg C14	LaIpmaSDLk	LYdVShSSMR	AKWAg-VAGA	TGYMLLYAPL	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE
Hs U1	LaIpmaSDLl	LYdVShSSMR	VKWDn-VpGA	SGYLLLYAPL	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE	TEGLAADEKE

var.

cons.

CAP Collagen alpha precursor

C12 Collagen type 12

FND Fibronectin type III domain

FN Fibronectin

TP Tenascin precursor

TC Tenascin-C

U1 Undulin 1

Bt Bovis taurus

Cf Canis familiaris

Ec Equus caballus

Ss Sus scrofa

Hs Homo sapiens

Oc Oryctolagus cuniculus

Xl Xenopus laevis

cow

dog

horse

pig

human

rabbit

African clawed frog

I L PGVD Y ITV A G S P

L I NTE VOL N R

AS R E

BOLD

lower case

non-conservative substitution

(charge reversal, change between hydrophobic

and charged, addition or removal of P)

position of non-conservative substitutions

Fig. 4

5/15



Fig. 5

6/15

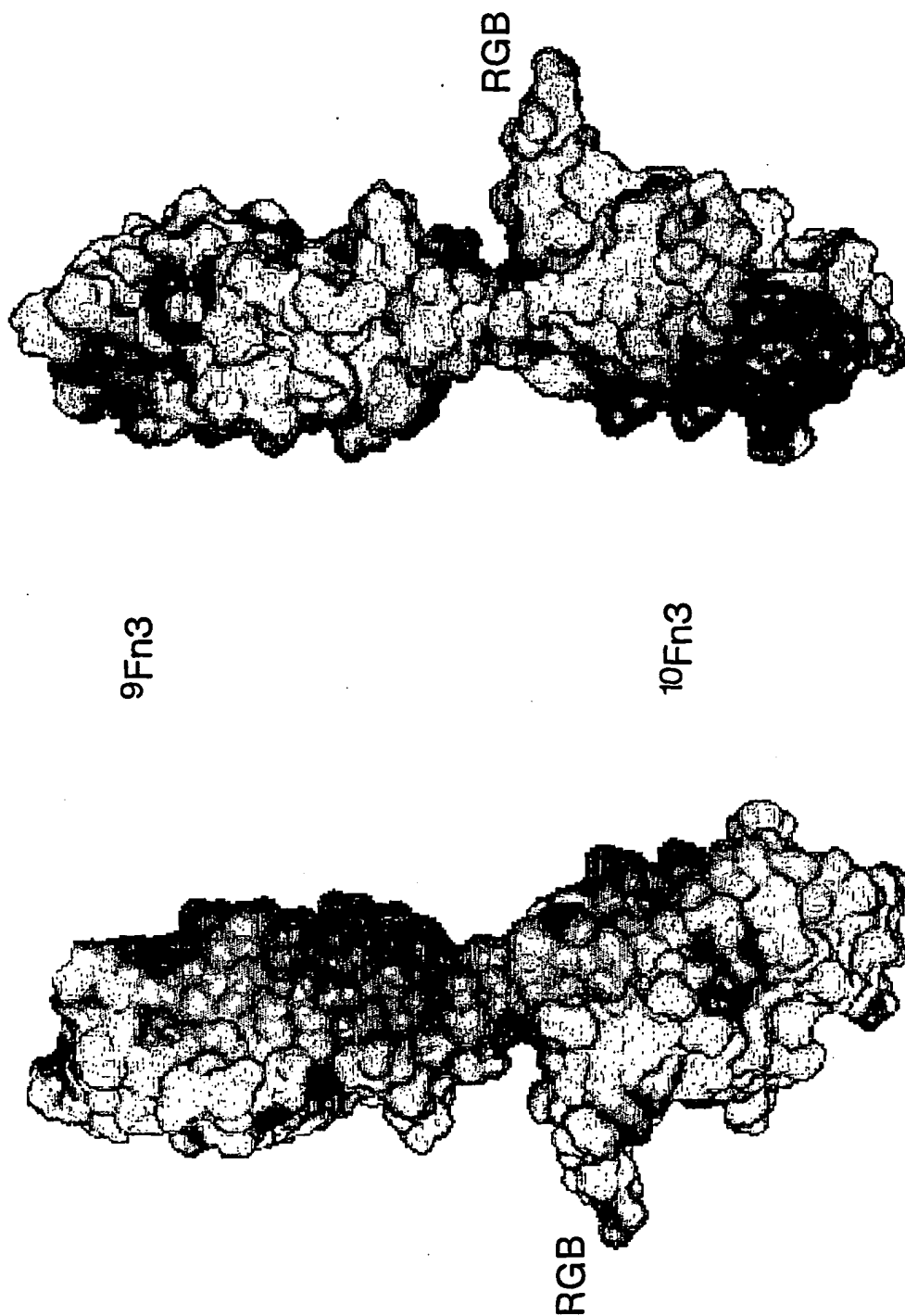


Fig. 6

7/15

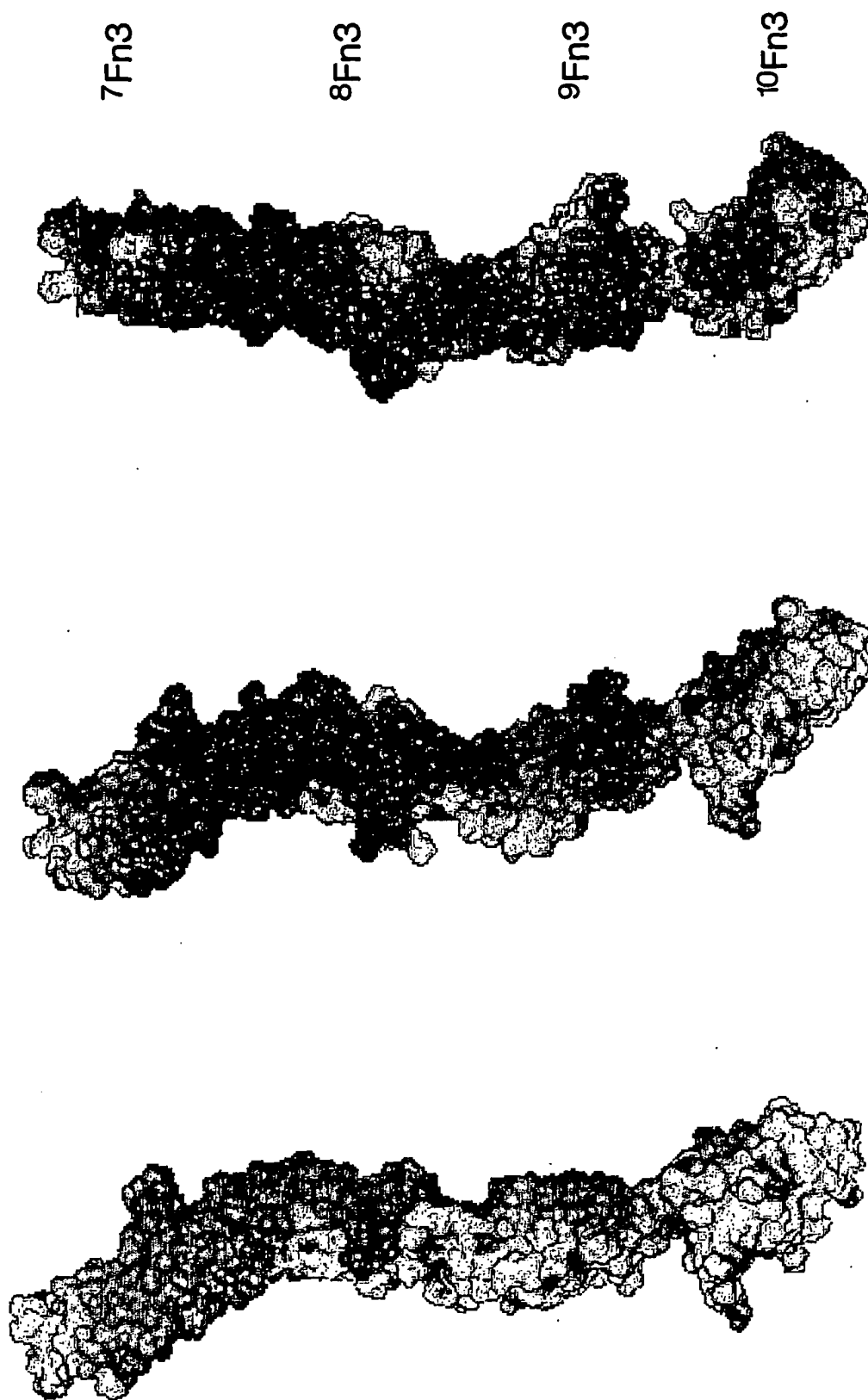


Fig. 7

8/15

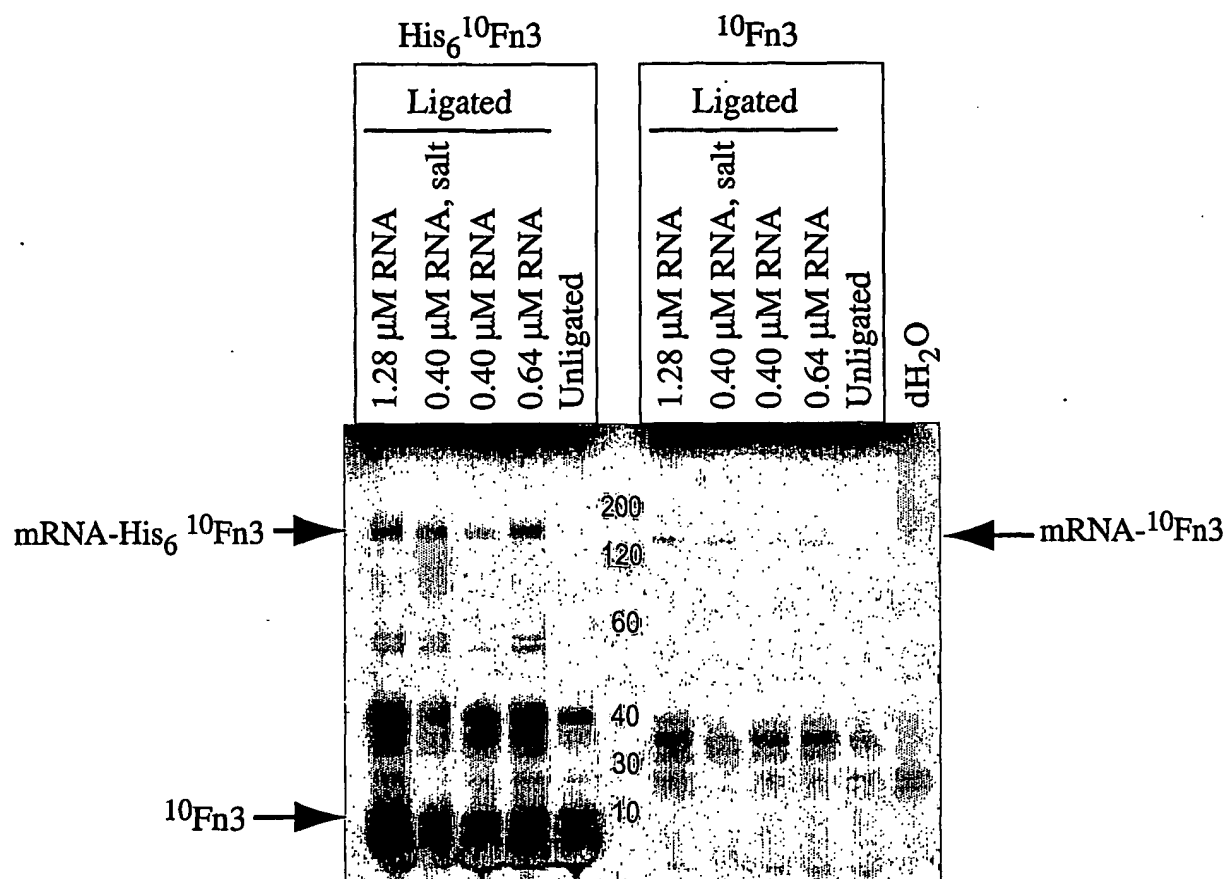
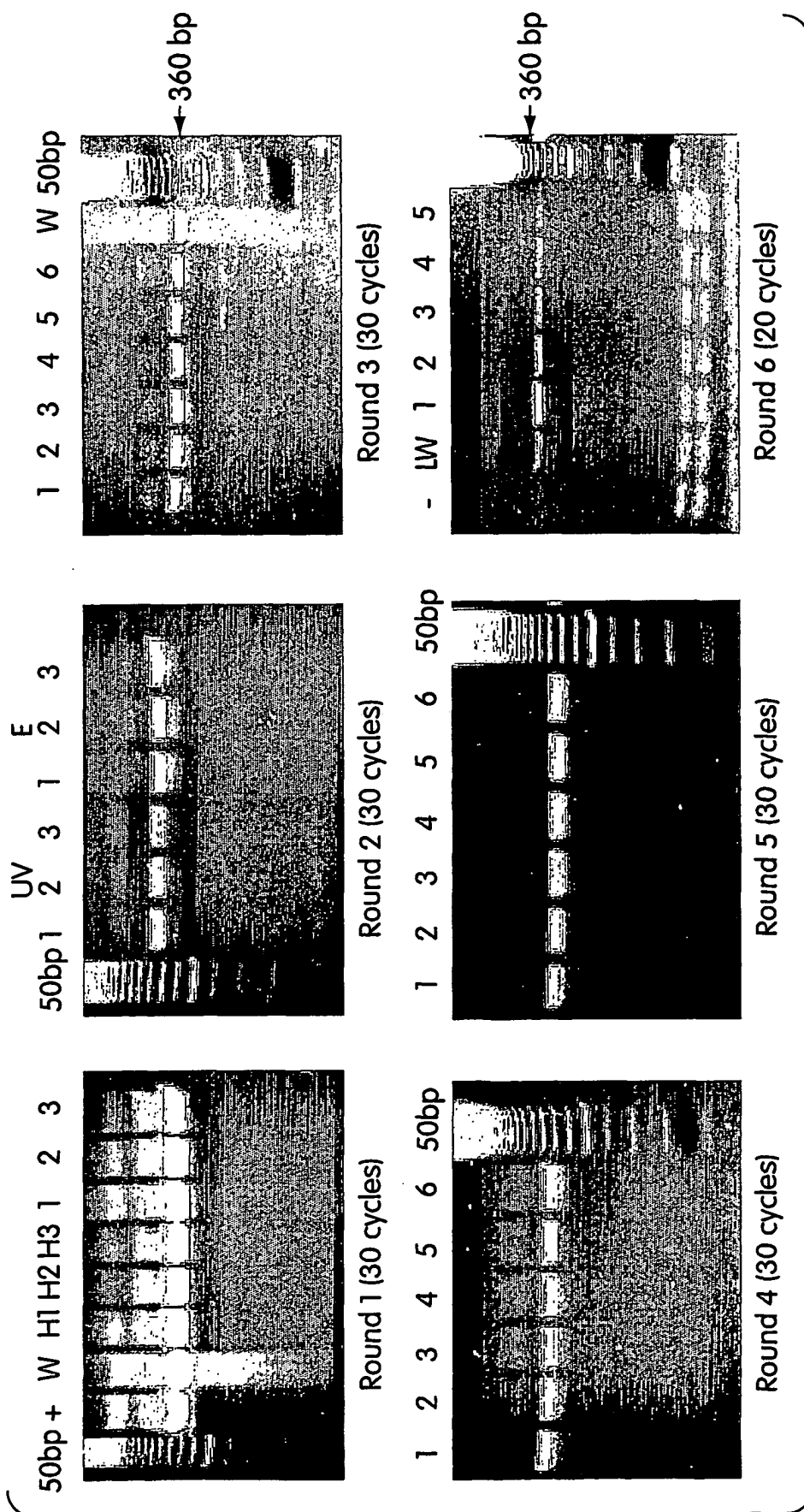


Fig. 8

9/15



10/15

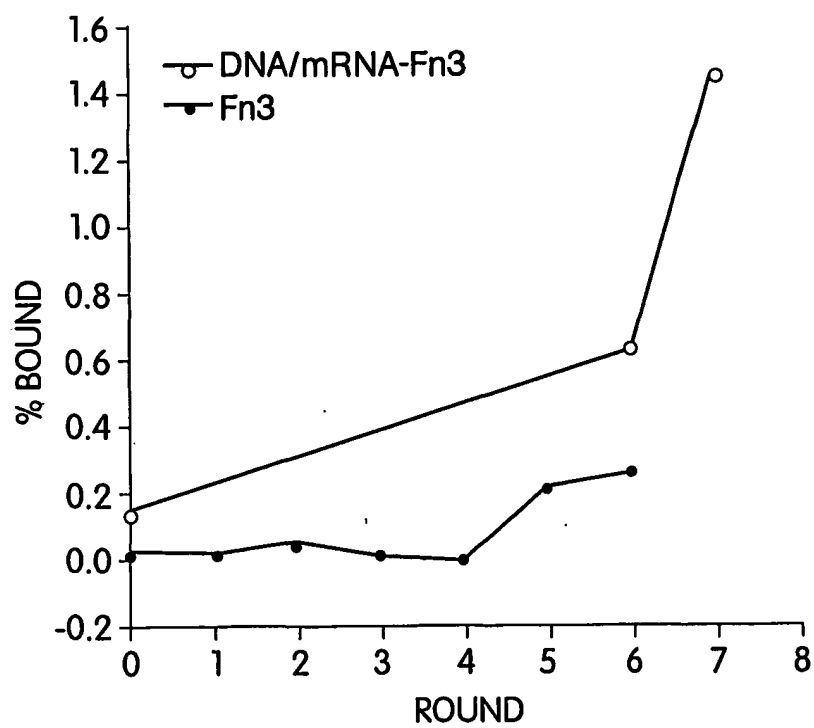
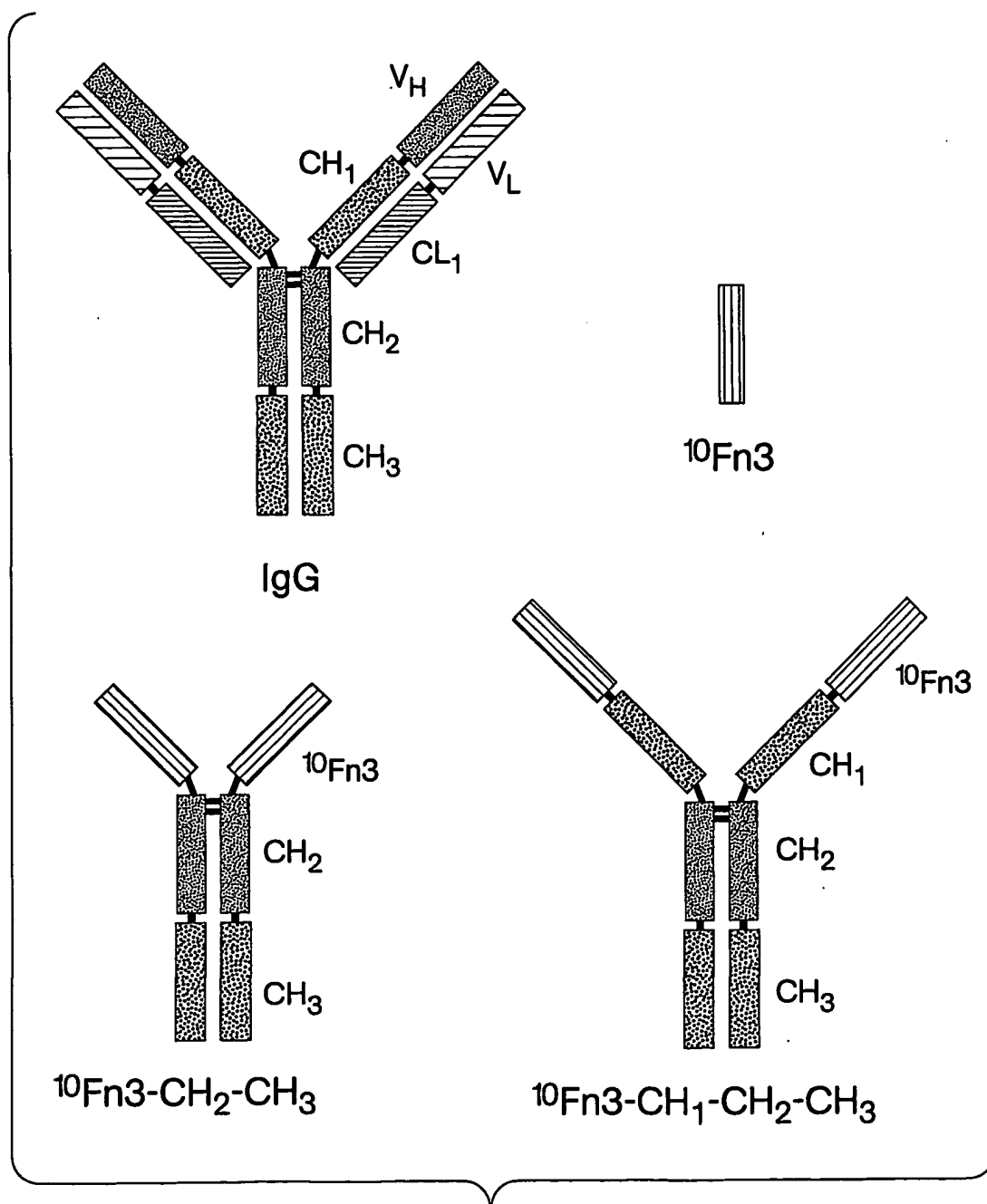


Fig. 10

11/15



12/15

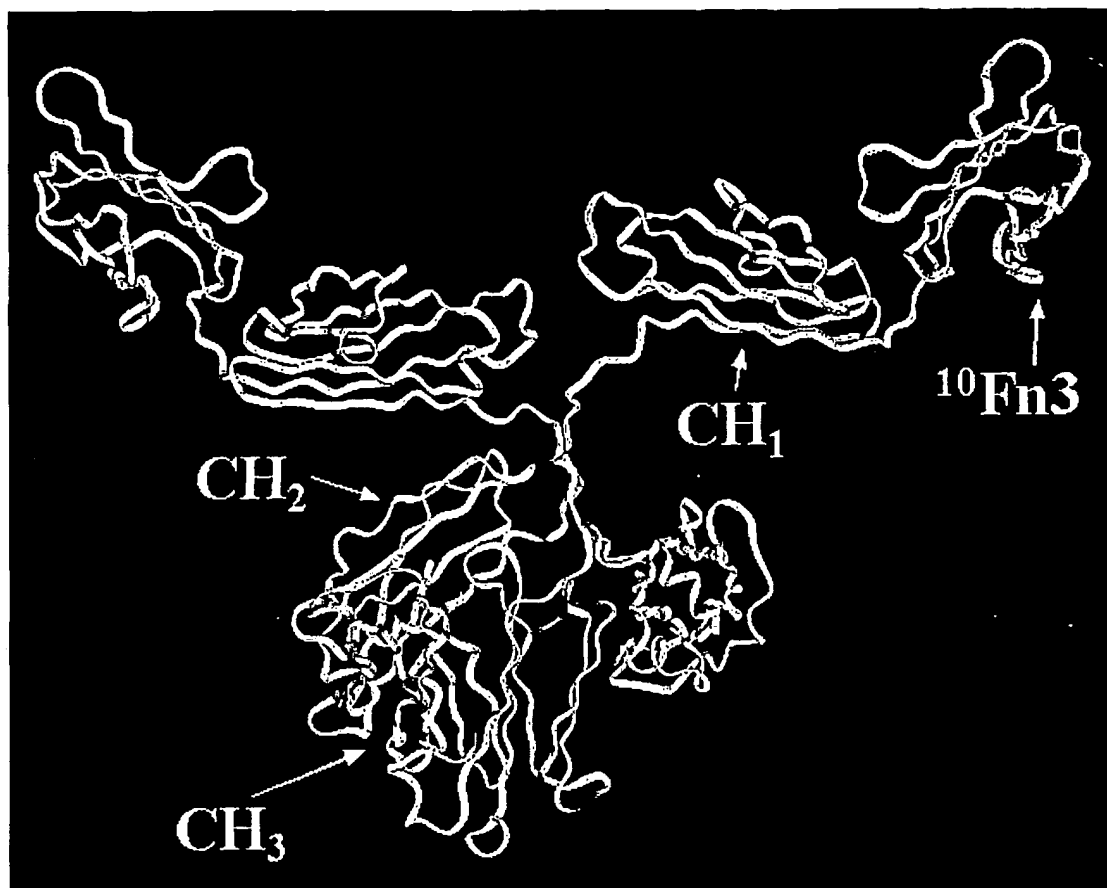
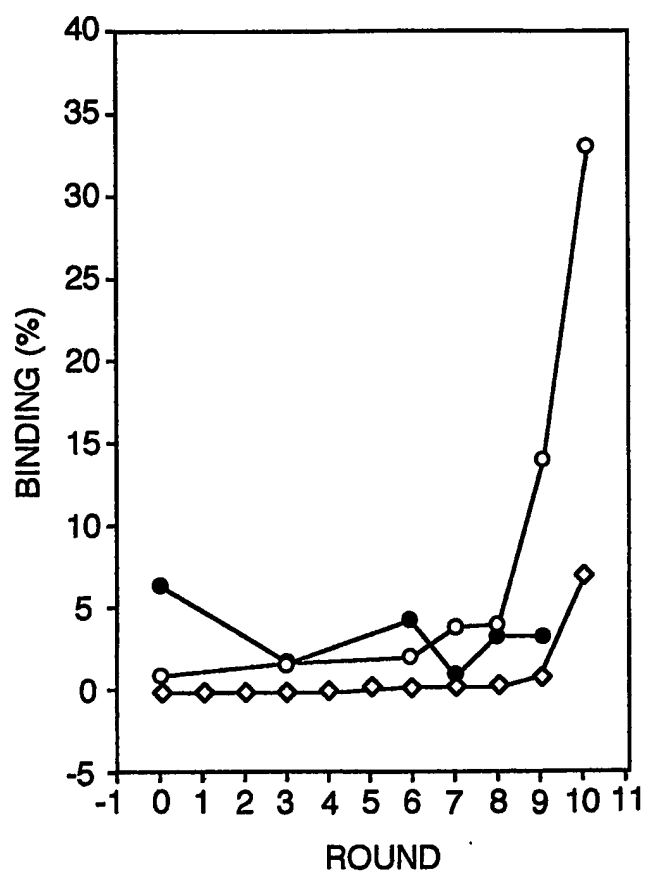


Fig. 12

13/15

FIG. 13



14/15

FIG. 14

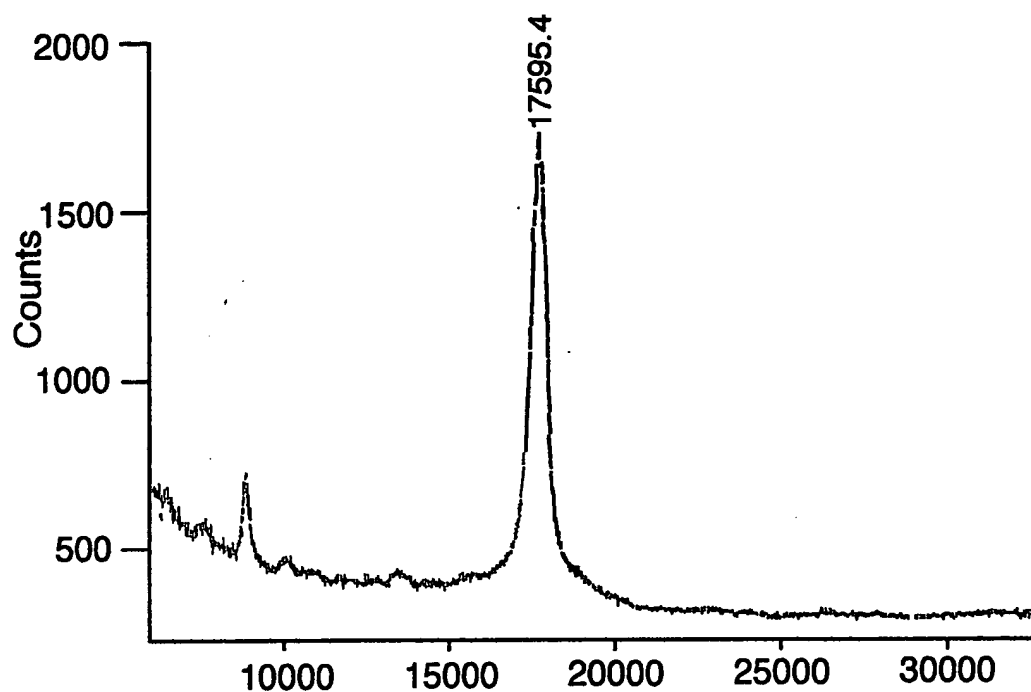
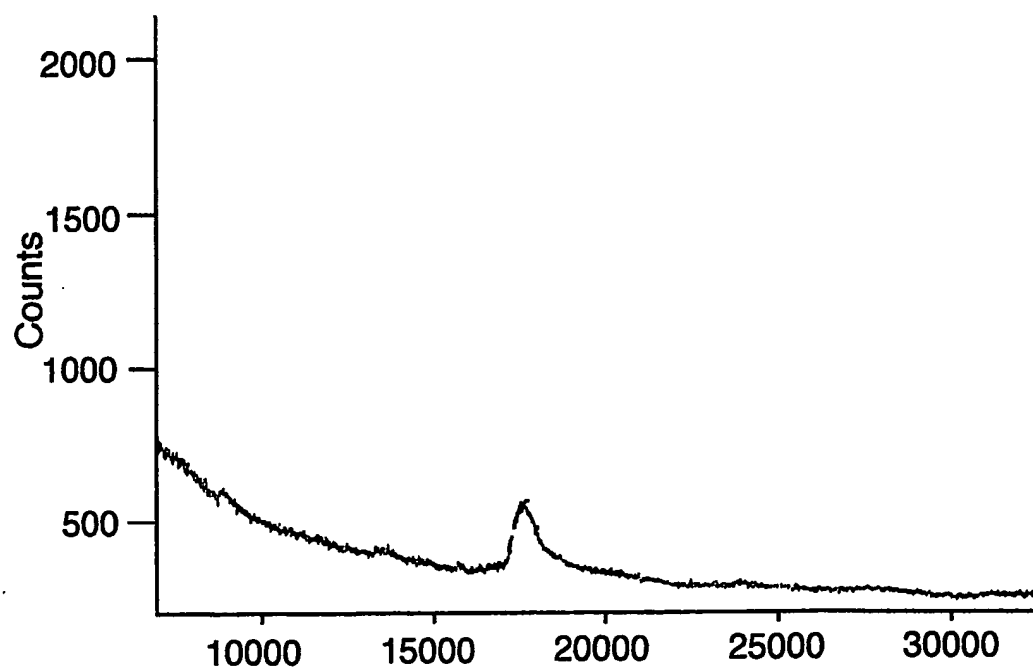


FIG. 15



15/15

FIG. 16

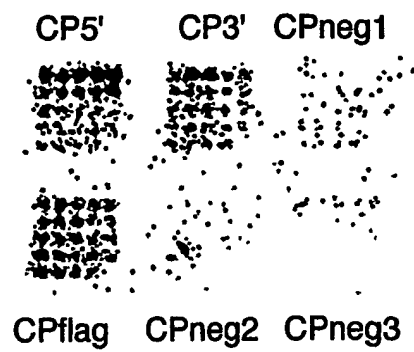
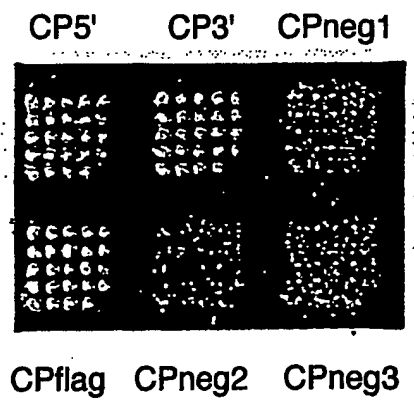


FIG. 17



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/06414

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/00, 1/70, 1/68; G01N 33/53

US CL : 435/4, 5, 6, 7.1, 69.1, 174

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/4, 5, 6, 7.1, 69.1, 174

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	KOIDE et al. The Fibronectin Type III Domain as a Scaffold for Novel Binding Proteins. J. Mol. Biol. 1998, Vol. 284, pp. 1141-1151, especially p. 1143, col. 2; p. 1144, Table 1; p. 1146, Col. 2.	1-6, 8-10, 15-21, 25-30 — 11-14, 17-19, 22-24, 31-34
X — Y	WO 98/56915 A2 (RESEARCH CORP TECH, INC.) 17 December 1998, entire document, especially p. 20, Table; p. 26, lines 5-6; p. 27, p. 32, Ex. 1; p. 35, Ex. 5; p. 40, Ex. 10.	1-6, 8-10, 15-16, 20-23, 25-30, 34 — 11-14, 17-19, 24, 31-33
X, P	WO 00/34784 A1 (PHYLOS, INC.) 15 June 2000, entire document.	1-34

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"I" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"G" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 29 JUNE 2001	Date of mailing of the international search report 07 AUG 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer HOLLY SCHNIZER <i>Dup Walling</i> Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/06414

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 98/31700 A1 (THE GENERAL HOSPITAL CORP.) 23 July 1998, entire document.	11-14, 24, 31-34
A	NYGREN et al. Scaffolds for engineering novel binding sites in proteins. Curr. Opin. Struct. Biol. 1997, Vol. 7, pp. 463-469.	1-34

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/06414

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

STN (bioscience database), EAST (all databases), search terms: fibronectin, module 10, 10fn3, scaffold, novel binding partners, combinatorial, phage display, evolve? novel binding, randomized loop.